JGR Biogeosciences

RESEARCH ARTICLE
10.1029/2021JG006743

Special Section:
Advances in scaling and modeling of land-atmosphere interactions

Challenges in Scaling Up Greenhouse Gas Fluxes: Experience From the UK Greenhouse Gas Emissions and Feedbacks Program

Peter Levy[1], Robert Clement[2], Nick Cowan[1], Ben Keane[3], Vasilis Myrgiotis[2], Marcel van Oijen[1], T. Luke Smallman[2], Sylvia Toet[3], and Mathew Williams[2]

[1]Centre of Ecology and Hydrology, Penicuik, UK, [2]University of Edinburgh, School of GeoSciences and NCEO, Edinburgh, UK, [3]University of York, York, UK

Key Points:

- We reviewed some of the challenges in accurately estimating fluxes of GHGs at a national scale
- Uncertainty arises from imperfectly-known models, parameters, and inputs used in the extrapolation
- Bayesian principles allow us to quantify this whilst combining information sources in a coherent way

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

P. Levy,
plevy@ceh.ac.uk

Citation:

Levy, P., Clement, R., Cowan, N., Keane, B., Myrgiotis, V., van Oijen, M., et al. (2022). Challenges in scaling up greenhouse gas fluxes: Experience from the UK greenhouse gas emissions and feedbacks program. *Journal of Geophysical Research: Biogeosciences*, *127*, e2021JG006743. https://doi.org/10.1029/2021JG006743

Received 2 DEC 2021
Accepted 28 APR 2022

Author Contributions:

Conceptualization: Peter Levy, Marcel van Oijen, Mathew Williams
Formal analysis: Peter Levy, Mathew Williams
Investigation: Robert Clement, Nick Cowan, Ben Keane, Vasilis Myrgiotis, Marcel van Oijen, T. Luke Smallman, Sylvia Toet
Methodology: Peter Levy, Marcel van Oijen

Abstract The role of greenhouse gases (GHGs) in global climate change is now well recognized and there is a clear need to measure emissions and verify the efficacy of mitigation measures. To this end, reliable estimates are needed of the GHG balance at the national scale and over long time periods, but these estimates are difficult to make accurately. Because measurement techniques are generally restricted to relatively small spatial and temporal scales, there is a fundamental problem in translating these into long-term estimates on a regional scale. The key challenge lies in spatial and temporal upscaling of short-term, point observations to estimate large-scale annual totals, and quantify the uncertainty associated with this upscaling. Here, we review some approaches to this problem and synthesize the work in the recent UK Greenhouse Gas Emissions and Feedbacks Program, which was designed to identify and address these challenges. Approaches to the scaling problem included: instrumentation developments which mean that near-continuous data sets can be produced with larger spatial coverage; geostatistical methods which address the problem of extrapolating to larger domains, using spatial information in the data; more rigorous statistical methods which characterize the uncertainty in extrapolating to longer time scales; analytical approaches to estimating model aggregation error; enhanced estimates of C flux measurement error; and novel uses of remote sensing data to calibrate process models for generating probabilistic regional C flux estimates.

Plain Language Summary Greenhouse gases cause climate change, and we need to know how much is emitted each year across the globe. As well as coming from burning fossil fuels, plants and soil also take up and emit these gases, and we need to be able to quantify this in order to understand how best to tackle climate change. However, we can only measure these emissions over very small areas, at only a few locations, and for relatively short periods of time. Extrapolating from these measurements to a whole country introduces several uncertainties which are often largely ignored. Here, we examine progress in tackling this problem and focus on better statistical methods to properly identify and account for the errors that are introduced by the large change in scale. Another focus is the development of instrumentation that can measure the gas emissions over larger scales and run continuously. Earth observation from satellites provides a promising source of data for the future, but cannot yet provide direct measurements of gas emissions. The so-called "Bayesian" approach to modeling, which allows us to relate observations to previous knowledge via the theory of conditional probability, provides us with a coherent method for combining data from different sources, accounting for their uncertainties, and propagating this through to the uncertainties associated with predictions of national-scale fluxes.

1. Introduction

The role of greenhouse gases (GHGs) in causing global climate change is now well recognized (IPCC, 2013). Emissions of GHGs from terrestrial ecosystems play an important part in this, and the potential for feedback within the climate system which amplify the emissions of GHGs from natural ecosystems is substantial. Accurate estimates are therefore needed of the GHG balance of the land surface at regional and national scales, and over long time periods, if we are to understand the key driver of global change. Because of the large scale involved, in relation to the scale at which we can make observations, this presents a major challenge that spans the domains of biogeochemistry, ecology, remote sensing, and atmospheric science.

Writing – original draft: Peter Levy, Mathew Williams
Writing – review & editing: Peter Levy, Robert Clement, Nick Cowan, Ben Keane, Vasilis Myrgiotis, Marcel van Oijen, T. Luke Smallman, Sylvia Toet, Mathew Williams

For directly measuring GHG fluxes, we have two approaches available, based on either enclosing a small area within a chamber and monitoring the change in GHG concentration, or based on micrometeorological measurements of GHG concentration and turbulence the near the surface (see the section below). However, both of these operate at scales much smaller than the spatial scale of interest - that of a region, nation, or the whole globe. This means that we need to use a model to predict the large-scale flux. The fundamental upscaling issue is that we are forced to rely on predictions from a model which cannot be parameterized or tested at the true scale of interest. If we introduce some generic notation, we can consider this as three inter-related problems. We need to predict y, the large-scale GHG flux, based on parameters θ derived at a small scale and input variables x estimated over the large-scale domain:

$$y = f(\theta, x). \tag{1}$$

First, the parameters θ are only inferred from a very small subset of the conditions prevailing over the whole domain. Because of basic sampling error, there is uncertainty in the parameter estimates. Second, there is also usually considerable uncertainty in the values of the inputs x over the large-scale domain. GHG fluxes from an ecosystem depend upon such things as incident radiation, leaf area index, soil aerobic status, soil microbial populations, and time elapsed since disturbance events. None of these is easily measured over a wide region, and we inevitably rely on some proxy or modeled estimate, and this introduces uncertainty in the true value of x over the whole domain. Third, the model f is commonly non-linear, which complicates the upscaling procedure. The goal for science in this field is to quantify and reduce the uncertainty associated with parameters θ, input variables x, and model f in making the jump between the small scale of measurement and the large scale of prediction, which we illustrate in Figure 1. Progress here is necessary if we are to estimate the large-scale GHG balance accurately (Leip et al., 2018), and to demonstrate the efficacy of mitigation policies (Gifford, 1994; Smith & Smith, 2004; Smith et al., 2008).

In this paper, we review the challenges in upscaling small-scale GHG flux measurements to produce national-scale estimates. The UK recently developed a novel, multi-disciplinary Greenhouse Gas Emissions and Feedbacks program (GHGEF) to identify and tackle some of these challenges, and our examples come from this program. The overall aim was to improve the quantification of uncertainty where it arises in the upscaling process and to reduce this uncertainty by improvements to instrumentation, measurement methods, or modeling procedures. Specifically, we focused on five challenges that focus on components of the problem, illustrated in Figure 1:

1. **Quantifying uncertainty in spatial upscaling of chamber fluxes to field scale.** Chamber measurements sample only a very small area, even in relation to a single agricultural field. The challenge is to quantify the mean and uncertainty in the estimate of the field-scale mean flux
2. **Quantifying uncertainty in temporal upscaling of chamber fluxes to annual scale.** Similarly, chamber measurements typically sample only during a very few hours, in relation to the total flux over a year. The challenge here is to quantify the annual cumulative emission and its uncertainty, based on a sparse and spatially variable sample set
3. **Reducing uncertainty in spatial and temporal upscaling of chamber fluxes** via **improved instrumentation.** An alternative approach to both of the above is to develop new measurement systems which can provide better spatial and temporal coverage
4. **Quantifying uncertainty in eddy covariance measurements of field scale fluxes.** Eddy covariance systems are expensive and complex to run, so are rarely operated with any replication. It is therefore usually very difficult to estimate the systematic and random errors associated with these measurements. Here we evaluate five co-located eddy flux systems to determine the measurement error on net exchanges of CO_2 at both instantaneous times and daily scales
5. **Quantifying aggregation error in spatial upscaling.** When non-linear models are parameterized at a small scale but applied at a larger scale, the results will generally be in error wherever small-scale heterogeneity is not accounted for. A further challenge is to estimate and account for this kind of error. We evaluated this in the context of national-scale GHG flux estimates in the UK

After outlining the basic approach to measuring GHG fluxes, we address each of these challenges in turn, with reference to specific analyses for a range of GHGs. For clarity, each section provides its own methods, results, and discussion. We then conclude with a synthesis of the findings from the individual studies. Our synthesis allows us to assess the advantages and limitations of current research, and to make suggestions for the development

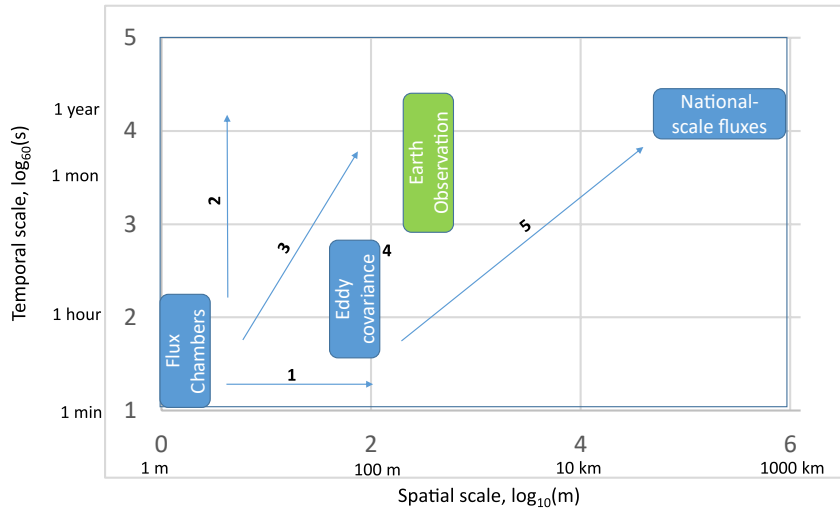

Figure 1. Diagram illustrating the spatial and temporal scales of the upscaling problem, from measurements using chambers and eddy covariance which cover small scales, to national-scale annual fluxes. The bars show the typical resolution of the different measurement techniques. The numbers respond to the challenges which we identify in tackling this problem. Earth observation represents a rapidly expanding source of data with wide coverage and increasingly fine resolution.

of new studies and approaches necessary to make better inferences about GHG fluxes at regional, national and global scales.

2. Measurement Methods for GHG Fluxes

For directly observing GHG fluxes, we have two broad techniques available: chamber-based and micrometeorological. In the former, part of the plant or soil surface is enclosed in a gas-tight chamber, and the flux is inferred from measurements of the mixing ratio. In the case of static (non-steady-state) chambers, the mixing ratio is measured on a sequence of gas samples extracted from the chamber over a short time period. From mass balance, the mixing ratio within the chamber is predicted to follow:

$$\chi = \chi_0 + \frac{F}{h\rho}\mathrm{d}t \tag{2}$$

where χ_0 is the initial mixing ratio of a GHG, h is the height of the chamber, ρ is the molar density of dry air, and $\mathrm{d}t$ is the time increment since enclosure. We thus have an inverse problem, which can be rearranged to estimate the flux as:

$$F = \frac{\mathrm{d}\chi}{\mathrm{d}t_0}h\rho \tag{3}$$

where $\mathrm{d}\chi/\mathrm{d}t_0$ is the initial rate of change in the mixing ratio. As an approximation, we can assume linearity in $\mathrm{d}\chi/\mathrm{d}t$, and solve for $\mathrm{d}\chi/\mathrm{d}t_0$ using linear regression. If we account for the non-linearity of diffusion into the chamber, we have to apply non-linear regression, optimization methods, and potentially, complex 2-D diffusion models (Livingston et al., 2006; Pedersen et al., 2010; Sahoo & Mayya, 2010; Levy et al., 2011). Because part of the ecosystem has to be physically enclosed, the spatial scale of these measurements is necessarily restricted, typically to 0.1 m^2 and rarely more than 1 m^2. Similarly, the temporal scale of measurements is restricted because the physical enclosure changes the environment within - the emitted gas concentrations build up, and the effect of wind and rain is removed.

Micrometeorological techniques make use of measurements in the atmosphere near the surface. Historically, these were based on measuring the gradients in the wind and GHG mixing ratios and making some assumptions about the turbulent transport. With the advent of fast-response infra-red analyzers for CO_2, and more recently for CH_4 and N_2O based on QCL or CRD laser absorption spectroscopy, the eddy covariance method has become the default approach (Haszpra et al., 2018; Kroon et al., 2010; Mammarella et al., 2010). If we can assume

stationarity and horizontal homogeneity, it follows from a mass balance that we can equate the surface flux to the eddy covariance term that is,:

$$F = \langle w'\chi' \rangle \langle \rho \rangle \tag{4}$$

where w' and χ' represent the instantaneous deviations from the means, and angled brackets denote time-averaged means. To measure this term accurately, we need high frequency (10–20 Hz) measurements of the vertical windspeed w and χ at an appropriate height above the surface. Various corrections are required to account for the frequency response of the measurement system, non-zero vertical windspeed, deviations from stationarity, and density fluctuations (Aubinet et al., 2012; Lee et al., 2006). The advantage of the approach is that it measures the integrated surface flux over an area much larger than a chamber, typically several hundred square meters (of the order of a small agricultural field), and can run near-continuously.

Eddy covariance systems can also be mounted on aircraft, to measure fluxes over a wider area from a very mobile platform. This provides a means to estimate the GHG fluxes over a region, and this has been used successfully in combination with tower-based measurements elsewhere in Europe and North America (e.g., Desjardins et al., 1997; Gioli et al., 2004). However, this type of measurement is difficult in the UK, partly because of civil aviation restrictions, and partly because of the fine scale of heterogeneity in the landscape. The logistics of finding areas with suitable topography, where the landscape units are large enough to be interpretable, and where permission to fly at low altitude can be obtained, makes it very problematic. Alternative airborne approaches, based on inverse atmospheric modeling of GHG mixing ratios observed from aircraft (Pitt et al., 2016; Polson et al., 2011) have shown some success, but have been limited to short campaigns, and are not considered further here.

3. Spatial Upscaling of Chamber Fluxes to Field Scale

As described above, chambers used to measure gas fluxes typically have small dimensions (<1 m^2), several orders of magnitude smaller than the domains, such as agricultural fields, that we want to make inferences about, so the potential for sampling error is large. That is, the naïve sample mean of the chambers may deviate substantially from the true mean for the field. We want to improve this estimate and quantify the associated uncertainty in extrapolating the field mean. This is a common problem in the area of geostatistics, where observations are only available at point locations, but predictions are required over a larger spatial domain. The classical geostatistical approach to this problem is to represent the spatial domain as a grid of discrete cells (a "raster") and to use kriging to predict the values at all the unobserved locations in this grid. Kriging and its terminology originated in the mining industry but is now a widespread and generally applicable technique for extrapolation problems. In essence, it is a form of weighted local averaging, where the estimates of values at unrecorded places are weighted averages of the observations. The kriging weights are calculated on the basis of the semivariogram, which quantifies the form of the increasing variance (or decreasing covariance) between pairs of points as the distance between them increases. Various mathematical models are used to describe the covariance as a function of this distance; the Matérn model is a common choice for its flexibility (Pardo-Iguzquiza & Chica-Olmo, 2008) in which the covariance is given by

$$C_\nu(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{d}{\phi}\right)^\nu K_\nu \left(\frac{d}{\phi}\right) \tag{5}$$

where σ^2 is the variance, ϕ is a range or scale parameter, ν is a shape parameter, Γ is the gamma function, and K_ν is the modified Bessel function. Graphically, this shows the scale at which values are highly correlated, and how this changes with spatial scale. Prediction at a new location is based on all the observations, each weighted according to the degree of correlation at that distance predicted by the semivariogram. Kriging has been shown to be optimal in the sense that it provides estimates with minimum variance and without bias (in the long-term statistical sense) (Cressie, 1990). It is also often described as providing estimates of known variance, but this is only true if the form of the semivariogram is known with certainty; in real-world applications, this is never the case. Here, we use kriging to extrapolate chamber fluxes to the field scale, but for the purposes of characterizing the uncertainty correctly, we apply it in a Bayesian framework. In brief, this means we account for the uncertainty in the variogram model and represent each of the parameters as a probability distribution. Rather than assuming the variance is known, we calculate the posterior distribution of the parameters, given the observed data, and

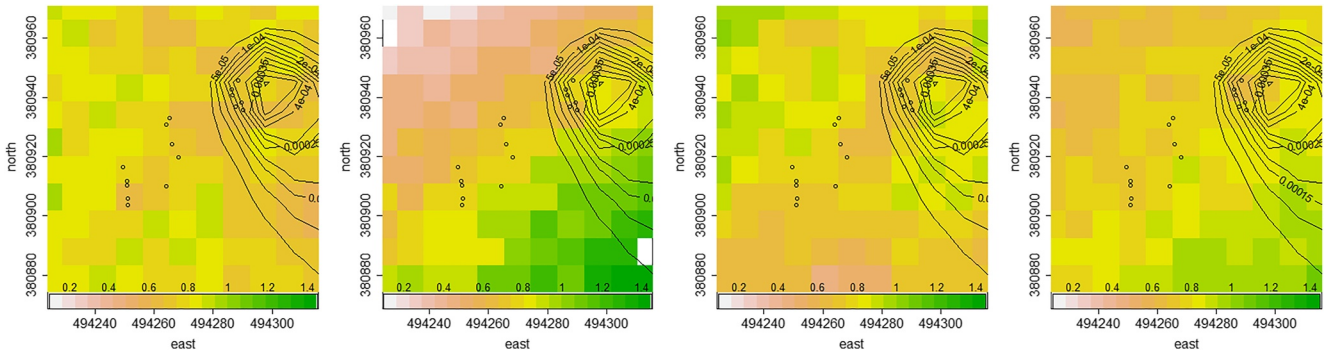

Figure 2. Four example realizations of the spatial pattern in N_2O flux from the posterior distribution for one sampling date (1 April 2015) in an arable field of oilseed rape in Lincolnshire, U.K. The color scale shows the N_2O flux in nmol m^{-2} s^{-1}. Points show the locations of the 16 flux measurement chambers. Contour lines show the footprint probabilities of the eddy covariance measurements over this chamber sampling period. On this occasion, the footprint was mainly to the south-east of the chamber locations. A range of quite different patterns are all plausible, given the chamber data.

sample many realizations of these to represent the uncertainty. Here, we applied the method to data from an arable field of oilseed rape in Lincolnshire, U.K., where fluxes of N_2O were measured by chambers (Keane et al., 2017). We used the Matérn model, with uniform (uninformative) priors for the σ, ϕ and ν parameters. This was implemented using the geoR package for the R statistical software (Ribeiro Jr et al., 2020).

We can attempt to test the success of this upscaling method because we can also measure at the field scale using eddy covariance. However, eddy covariance also does not directly give the field-scale domain mean, as its spatial sampling characteristics are affected by measurement height, surface roughness, wind speed, wind direction, sensible heat flux, and friction velocity: the so-called "flux footprint" (Leclerc & Foken, 2014; Rannik et al., 2012; Schmid & Oke, 1990; Schuepp et al., 1990). The footprint defines the relative contribution of each element of the surface area to the measured vertical flux, according to the advection-diffusion equation. This acts as a weighting function, such that some areas contribute strongly to the measured flux, and others do not at all. If the mean flux F of a scalar over a landscape represented by a discretized gridded domain with dimensions n_x by n_y at time t is given by:

$$\bar{F}_t = \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} F_{xyt} \frac{1}{n_x n_y} \tag{6}$$

where the overbar denotes spatial averaging, eddy covariance effectively measures a weighted mean, where the footprint provides the set of weights, ϕ, to give:

$$\hat{\bar{F}}_t = \sum_{x=1}^{n_x} \sum_{y=1}^{n_y} F_{xyt} \phi_{xyt} \tag{7}$$

where we use the $\widehat{\text{hat}}$-symbol to indicate that this is an estimator, not $\bar{F}_t$ itself.

So the appropriate way to upscale chamber measurements is to use Bayesian kriging to estimate F_{xyt} over the whole grid, and thereby the domain mean $\bar{F}_t$. However, to compared with eddy covariance, we need to apply the footprint weighting ϕ to estimate the flux that we would expect eddy covariance to measure, $\hat{\bar{F}}_t$.

Here, we could test this using eddy covariance measurements from the same field as the chamber measurements (Keane et al., 2017). The Bayesian kriging method provides a way of scaling chamber flux to the field scale, incorporating spatial pattern and its associated uncertainty. This prediction is weighted by the flux footprint (calculated here by the model of Kormann & Meixner [2001]) to produce our expectation of the flux measured by eddy covariance. This allowed us to compare one method with another, accounting for the difference in spatial sampling characteristics inherent in the two methods as best we can and account for the associated uncertainty properly.

Figure 2 shows four example realizations of the spatial pattern in N_2O flux from the posterior distribution for one chamber sampling date. This shows that a range of quite different patterns is all plausible, given the chamber data.

The footprint of the eddy covariance measurements over this chamber sampling period is superimposed as the contour lines, in relation to the 16 chamber locations. On this occasion, the footprint was mainly to the southeast of the chamber locations. This illustrates the difficulty in relating the two measurement techniques is that there is considerable uncertainty in inferring the N_2O flux within the footprint region, based on the 16 chamber locations we have available and the information on the spatial pattern which we extrapolate.

Figure S1 in Supporting Information S1 shows the chamber fluxes of N_2O upscaled by Bayesian kriging and weighted by the footprint probabilities ϕ, to give a valid comparison against the eddy covariance data. The results show that the upscaled values can deviate substantially from the naïve sample mean of the chambers. More importantly, the uncertainty in the mean estimated by Bayesian kriging is substantially larger than the conventional 95% CI in most cases. This is a source of uncertainty that is typically ignored, and this demonstrates the importance of representing spatial upscaling effects explicitly. This also indicates that we need to be careful in drawing conclusions from such measurements when interpreting the difference in means among experimental treatment plots. The magnitude of error will depend on the spatial pattern in the surface flux; in extreme situations, the sample mean could be quite poorly representative of the large-scale mean. In these cases, spatial upscaling clearly needs to be considered explicitly, and the Bayesian kriging described here provides a rigorous method to do this.

When comparing chambers and eddy covariance measurements, the difference in spatial sampling is usually ignored, and the arithmetic mean of chamber flux samples is compared with the value from eddy covariance for the corresponding time period. This ignores the fact that the flux footprint acts as a moving spatial filter, whereby the location and extent of the area that influences the measured flux changes each half-hour, according to wind speed, direction, etc. Spatial heterogeneity will always have some influence on eddy covariance flux measurements. In most cases, we choose our sites to be as homogeneous as possible, so we can hopefully ignore the problem. Where the landscape within the fetch can be classified into discrete classes, the contribution of each class to the measured flux can be characterized by the footprint and used as a weighting, and there are several examples of this approach (e.g., Morin et al., 2017; Sachs et al., 2010; Schrier-Uijl et al., 2010). However, this still assumes within-class homogeneity. A few studies have looked at inferring the spatial pattern, allowing for different surface classes and a random spatial effect (Levy et al., 2020), and representing continuously-varying spatial effects (Reuss-Schmidt et al., 2019). Our method here is novel in that it accounts for a continuously-varying spatial pattern, and furthermore, the distribution of possible continuously-varying spatial patterns that are consistent with the chamber data. This provides a rigorous way to compare chamber measurements so that they represent the same area that is sampled by eddy covariance, allowing for uncertainty in making the spatial extrapolation.

4. Temporal Upscaling of N_2O Fluxes to Annual Scale

There are considerable challenges in interpolating and extrapolating cumulative N_2O fluxes, based on relatively sparse and variable measurements from only a few time points. This leads to substantial uncertainty in the "emission factor" (EF or Ω, the total N_2O released as a percentage of the fertilizer nitrogen added) which is used in the national inventory. This is analogous to the problem of estimating the spatial mean from measurements at a limited number of locations but in the time domain. The method most commonly used in the literature to calculate cumulative N_2O fluxes is to interpolate and integrate using trapezoidal rule integration. However, this method is very sensitive to noise in the data (Smith et al., 2003), there is no straightforward way to quantify the uncertainty introduced or to extrapolate beyond the sample data, and it does not account for the typically lognormal spatial distribution of fluxes (Cowan et al., 2019). Here, we used two approaches to examine spatial and temporal upscaling: firstly, using a process-based model directly, and secondly, using Bayesian emulation of this model.

4.1. DNDC Model

DNDC (DeNitrification-DeComposition) is a process-based biogeochemical model (Haas et al., 2013), widely used to estimate agricultural soil N_2O emissions (Gilhespy et al., 2014). We parameterized the model using chamber measurements of N_2O fluxes, along with data on crop nitrogen content, soil mineral nitrogen, and soil moisture, from four experimental sites between 2010 and 2012, accounting for time lags between measured and simulated time-series (Myrgiotis et al., 2016). The model was applied across a 3,800 km^2 area of Scotland where >90% of its croplands are located. Spatial data on soil properties, crop coverage and weather, and UK-specific

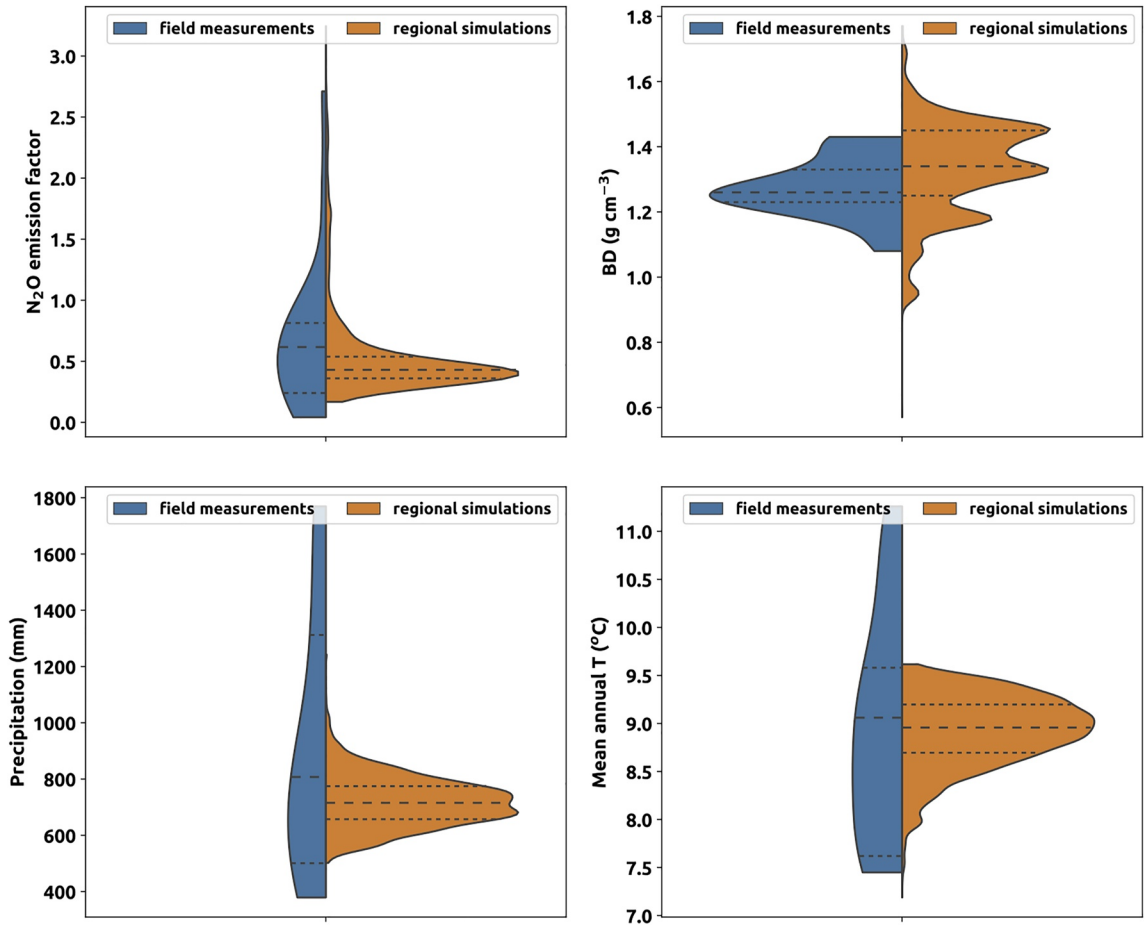

Figure 3. A comparison of distributions of emission factor Ω for N_2O generated from (i) field measurements collected across range of different management approaches, sites and seasons at locations in the UK and (ii) from upscaling estimates based on a model calibrated using the site level data, applied across eastern Scotland over several years. Also shown are the distributions of soil conditions (bulk density, BD) and weather conditions during measurement periods for all sites, and the conditions relevant to the modeled upscaling. The crops grown in the measured and simulated sites were winter wheat, winter barley, spring barley and winter oilseed rape.

crop calendars and fertilizer-use recommendations were used to create model inputs for 2011–2013 at a 1 km^2 resolutin (Myrgiotis et al., 2019).

The distribution of Ω predicted in the regional simulations was compared against measurements of Ω at the field sites. These had clear differences (Figure 3) in the range and shape of distributions. The wider range in measured Ω shows that there is more variability in the measured data than can be explained by the model. The model smooths this out, but there is considerable uncertainty that should be included in any upscaled predictions. Another factor is that the calibration data set came from sites with a different range of climatic conditions, compared to the Scottish arable region, so an exact like-for-like comparison is not possible. However, given that the inter-quartile range of the simulated Ω was around half that found in the measurements, we conclude that direct upscaling using the model leaves out substantial aleatoric uncertainty which should be represented in the national-scale estimate.

4.2. Meta-Model of N_2O Fluxes

To address this issue, we developed an emulator of the DNDC model with which we could apply Bayesian calibration to characterize the uncertainties in the spatial and temporal distribution of emissions. Following a fertilization event, the time course of N_2O flux is expected to rise to a peak, then decay exponentially. This pattern in time is reproduced by DNDC and similar models, and is well described very simply by the lognormal equation:

$$\mu_t = \frac{1}{\sqrt{2\pi}kt} e^{-(\log(t)-\Delta)^2/2k^2} N_{\text{in}}\Omega \tag{8}$$

where μ_t is the spatial mean of the N_2O flux at time t, Δ and k are analogs for the location and scale parameters, N_{in} is the nitrogen input, and Ω is the fraction of this nitrogen that is released as N_2O. Because the lognormal function integrates to unity at $t = \infty$, Ω is implicitly based on the total cumulative emission, rather than at an arbitrarily defined time. The symbol Δ can be interpreted as the natural logarithm of the delay between fertilizer application and peak flux; k is a decay rate term. This equation provides a simple meta-model that can be used to emulate the behavior of DNDC (and similar models), but includes the aleatoric uncertainty in the emission factor.

Because μ_t typically has a very skewed spatial distribution, there is a high probability of the sample means underestimating the true value; the problem increases as variance increases and the sample size decreases. Several approaches have been proposed as more efficient estimators of the location and scale of lognormal distributions, but none of these entirely solve the problem when σ is large and n is small, as is generally the case with flux measurements. In this study, we applied a Bayesian approach, using the Markov Chain Monte Carlo (MCMC) method with Gibbs sampling (Gelman et al., 2013). In this way, we estimated the parameters of the underlying distribution.

So, at time t following fertilisation, the mean flux is given by Equation 8, at which time the N_2O flux has a distribution

$$\begin{aligned} F &\sim \ln \mathcal{N}\left(\mu_{\log,t}, \sigma^2_{\log}\right) \\ \mu_{\log,t} &= \log\left(\mu_t\right) - 0.5\sigma^2_{\log} \end{aligned} \tag{9}$$

To obtain the cumulative flux at time t, we use the standard lognormal cumulative distribution function

$$F_{\text{cum}}, t = \Phi\left(\frac{\ln t - \Delta}{k}\right) N_{\text{in}}\Omega \tag{10}$$

where Φ is the cumulative distribution function of the standard Normal distribution. The model was encoded in the JAGS language (Plummer, 2016) which allows the model parameters to be estimated using Gibbs sampling in a Monte Carlo Markov chain algorithm. Parameters were estimated for a number of data sets from across the UK.

Priors for Δ and k were specified as Normal distributions based on the temporal patterns produced by the DNDC model (see below). A Normal distribution was also assigned to $\sigma_{\log}$, based on earlier data from various sites in the UK, mainly from Cowan et al. (2014, 2016). The prior distribution for $\sigma_{\log}$ was truncated at zero to exclude negative values. Ω was given a lognormal distribution, fitted to the data collation of Stehfest and Bouwman (2006) which included data on emission factors from all over the world.

Figure 4 shows the posterior distribution of cumulative fluxes calculated from a sample of the UK data sets described in Levy et al. (2017), expressed as the emission factor, Ω. Each data set represents a number of weeks after fertilizer was applied at a site, and N_2O fluxes were measured by chambers, as described by Levy et al. (2017). The uncertainty in the emission factor, expressed as the width of the posterior distribution is very variable, being narrowly defined in some cases (e.g., Dum 2012-10-16, EBS 2009-03-17), and very wide in other cases (e.g., EBS 2007-05-16, EBS 2008-06-18). Ω is generally in the range of 0 to 5%, but some events have substantially higher emission factors. The value estimated by the trapezoidal method is generally within the posterior distribution of the lognormal model but lacks any associated measure of uncertainty. The values from the trapezoidal method are usually lower. The emission factor is generally rather poorly constrained by flux chamber measurements, because of the difficulties of accurately estimating the mean of a lognormal distribution with large variance when n is small. The standard approach fails to capture this uncertainty. Our new approach performs well in that it appropriately quantifies the uncertainty, and removes some of the bias by accounting explicitly for the lognormal distribution.

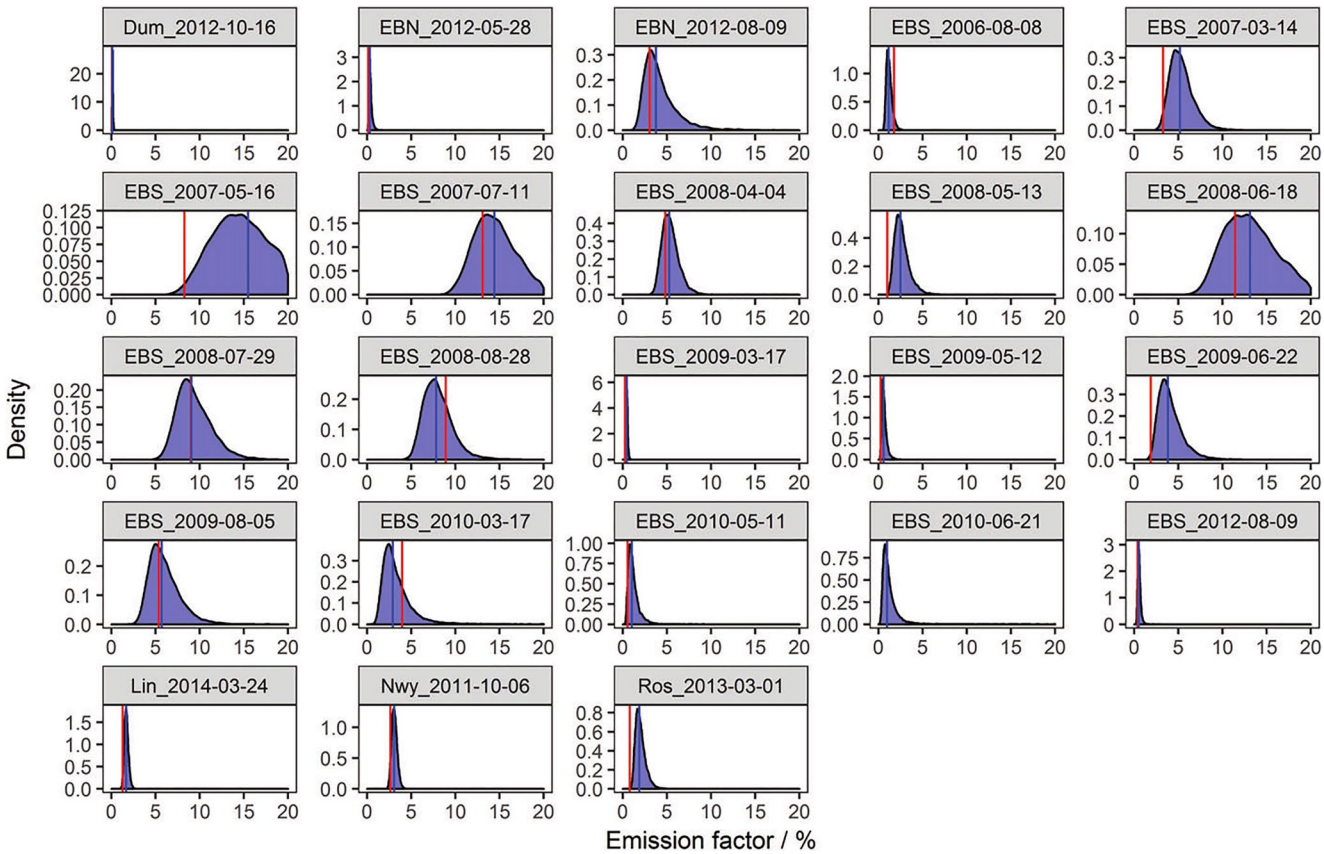

Figure 4. Posterior distribution of emission factor for each fertilizer application obtained by Bayesian estimation of the lognormal model. The blue vertical line shows the median of this distribution and the red vertical line shows the value calculated by the trapezoidal method. A lognormal distribution was used to provide the prior distribution of the emission factor. Panel headings show the site code (Dum, Dumfries; EBN, Easter Bush North; EBS, Easter Bush South; Lin, Lincolnshire; Nwy, North Wyke; Ros, Rosemaund) and date of fertilizer application, as described by Levy et al. (2017).

5. Reducing Uncertainty in Large-Scale Fluxes via Improved Instrumentation

The direct measurement of GHG fluxes has developed in tandem with the instrument technologies that allow GHG mixing ratios to be measured. Compared with CO_2, N_2O is less amenable to measurement by infra-red absorption: it is present at lower background concentrations; the typical fluxes are smaller relative to the background concentrations; and the infra-red absorption bands are narrower, making the technicalities of measurement more difficult. Until recently, N_2O was only accurately measurable by gas chromatography. Hence, for the most part, observations of N_2O fluxes are only available using static chamber methods (Hutchinson & Mosier, 1981; Matson & Harriss, 2009), which necessarily sample small areas (typically < 0.1 m^2) over short time periods, based on few (2–4) points. However, N_2O fluxes show wide variability, ranging over orders of magnitudes on small spatial scales, which is not predictable. This is attributed to two main causes: the variety of unobserved microbial controls on gas production, including their physiological activity and population dynamics; and the sensitivity of physical transport in the soil, which interacts with the measurement process (enclosure) in a complex way (Sahoo & Mayya, 2010; Xu et al., 2006). Fast-response sensors for N_2O have recently become available, making high-precision chamber measurements (Cowan et al., 2014) and the micrometeorological eddy covariance method (Kroon et al., 2010) feasible. The challenge here is to use these new sensors to develop continuous measurement systems for fluxes of N_2O. With continuous measurements, we overcome (some of) the need to interpolate and extrapolate in space and time, and thereby remove the large uncertainties this introduces. In the GHGEF program, we developed and applied two systems that provide near-continuous measurements of N_2O flux: a robotic auto-chamber system using a cavity ring-down spectroscopic instrument ("SkyLine"), and an eddy covariance system based on a quantum cascade laser (QCL) spectroscopy.

5.1. SkyLine

A detailed description of the system is available in Keane et al. (2017). To repeat this briefly, the SkyLine2D automated chamber system used a single, cylindrical chamber (internal diameter 40 cm, height 62 cm), suspended from a motorized trolley mounted on parallel horizontal ropes held above the crop by 2.5-m tall aluminum trellis arches (Figure S2 in Supporting Information S1). The trolley repeatedly traversed a transect across the crop of up to 40 m, enabling measurements at high spatial resolution (<1 m) across replicated manipulations of underlying variation in the landscape. At designated locations where collars were placed in the soil, the chamber automatically lowered and sealed on the collar to conduct a flux measurement. The base of the chamber was fitted with a rubber gasket which formed a gas-tight seal when dropped on the flange of the landing base. Guides around the chamber bases ensured the chamber landed accurately. A vent was fitted into the chamber to minimize pressure differences between the chamber and the external atmosphere (after Xu et al. [2006]). The chamber operated as a non-steady state dynamic system, with headspace gas being circulated between the chamber and a cavity ring-down spectroscopic (CRDS) analyzer for N_2O (LGR isotopic N_2O analyzer, Los Gatos Research, CA, USA) housed in an enclosed shed at one end of the SkyLine2D apparatus (Figure S2 in Supporting Information S1). The CRDS analyzer operated at 1 Hz, giving a precise measurement of the rise in mixing ratio within the chamber, allowing a relatively short chamber closure (ca. 5 min), thus minimizing the time during which the underlying plants and soil are isolated from ambient conditions.

5.2. QCL Eddy Covariance System

The system used a continuous-wave quantum cascade laser (QCL) absorption spectrometer (CW-QC-TILDAS-76-CS, Aerodyne Research Inc., Billerica, MA, USA), with an ultra-sonic anemometer (WindMaster Pro 3-axis, Gill, Lymington, UK) to measure fluctuations in 3-D wind components at a frequency of 20 Hz. The QCL was fitted with a laser capable of measuring N_2O with a precision of 0.03 ppb, together with H_2O and either CO_2 or CO, using absorption features at wavenumbers 12 and 22 m^{-1}. Internal software fits the observed spectra to a template of known spectral line profiles from the HITRAN (HIgh-resolution TRANsmission) molecular spectroscopic database. Absolute gas concentrations can then be calculated from the strength of the absorption line measured, the temperature, pressure, and path length. A vacuum pump (Triscroll 600, Agilent Technologies, US) was used to draw air through the inlet and instrument with a flow rate of approximately 14 l min^{-1}. Data from the sonic anemometer and QCL was logged in tandem using a custom program written in LabView (National Instruments, TX, USA).

Fluxes were calculated over 30-min intervals using the EddyPro software (Version 6.2.1, Li-COR, Lincoln, NE, U.S.A.), based on the covariance between gas concentration and vertical wind speed. For flux data taken with a low signal-to-noise ratio, time lag identification by maximization of the cross covariance introduces systematic biases (Langford et al., 2015). Here, we investigated methods for estimating the time lag, based on either maximizing the time lag over a longer time window, or using the time lag established for CO_2. Both CO_2 and N_2O share the same path in the sample line to the measurement cell, and would be expected to travel at the same rate. Fluxes of CO_2 are an order of magnitude larger, so the time lag which gives the maximum covariance is usually clearly defined within each half-hour and should be equally applicable to N_2O, and not subject to the systematic error described by Langford et al. (2015).

Various different lasers can be installed in the QCL system for measuring different gases, so measurements of CO_2 are not always available. The time lag is a critical uncertainty in the N_2O flux estimation, so it is important that we can establish a basis for estimating this in the absence of a CO_2 signal. Although estimating the time lag by covariance maximization introduces biases when applied at short timescales, the problems lessen as this timescale increases. We investigated this by systematically varying the time window over which covariance was maximized, so as to explore the range of timescales over which the time lag was identifiable, stable, and in agreement with the time lag identified by the CO_2 signal. Based on the results, we opted for a six-hour window (i.e., 3 hr before and 3 hr after) as the best compromise between the various competing issues. For each 30-min period, the time lag was based on maximizing covariance in this moving window. Standard corrections were made in the flux calculation, following Moncrieff et al. (1997), Ibrom et al. (2007), and Burba (2013). Random uncertainty was estimated by the method of Finkelstein and Sims (2001).

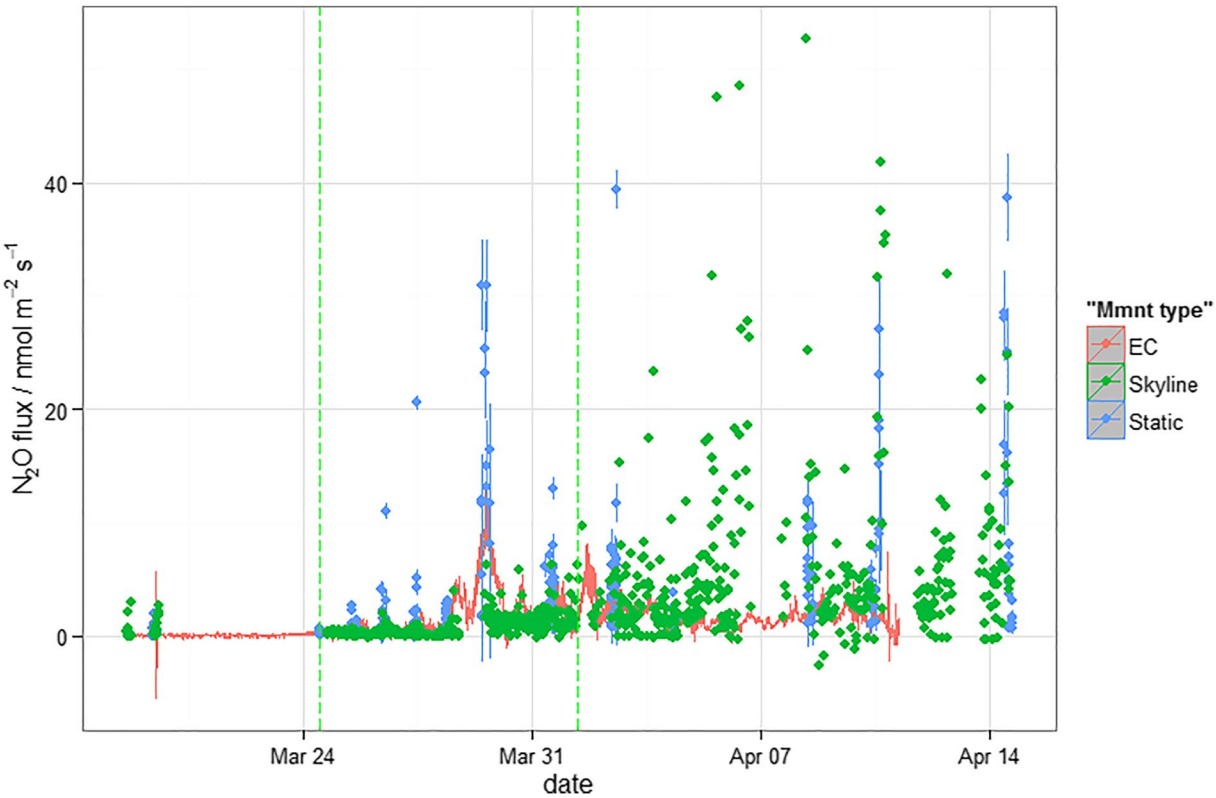

Figure 5. Time course of N_2O fluxes at the Lincolnshire field site, following fertilization events (dotted vertical lines) as measured by different types of the measurement system (Mmnt type): eddy covariance (EC), static chambers, or the SkyLine2D system.

Figure 5 shows a near-continuous time series of measurements of N_2O flux over a whole month by both automated SkyLine chamber and eddy covariance methods. Conventionally, only infrequent static chamber measurements (blue symbols) would be available. Given the variability in the data - irregular peaks in time and spatial variability between chambers - the near-continuous measurements provide a much more accurate estimate of the cumulative emission over the period following fertilizer application, and thereby the emission factor. Measurement techniques such as these, with better data coverage, are clearly needed to enable appropriate temporal upscaling so that longer-term means and cumulative emissions can be estimated accurately from the observations.

6. Quantifying Uncertainty in Eddy Covariance Measurements of Field Scale Fluxes

As described above, eddy covariance (EC) is one of the key techniques for measuring GHG fluxes, but is subject to instrument noise, uncertainties in the processing steps, and micrometeorological conditions required to meet the underlying assumptions. The measurement uncertainty for eddy covariance is thus challenging to quantify. Due to their expense, EC systems are usually deployed singly, so there is a lack of replication – the most obvious means by which this can be estimated. By using two EC systems 800 m apart, Hollinger and Richardson (2005) were able to quantify the measurement difference (uncertainty) in flux calculations. However, the tower separation was large enough so that footprint regions did not overlap, so the comparison is potentially confounded by the difference in the surface flux in the two footprint regions. Hollinger and Richardson proposed a time–for–space substitution to allow uncertainty calculations from a single EC system, the so-called successive days method, and showed it has utility.

Here for the first time, we present a comparison of NEE flux estimation from five co-located flux systems, all sampling the same pasture in southwest Scotland (Crichton Research Farm, Dumfries). The low stature of the vegetation allowed the EC systems to be set up within 10 m of each other, each sampling at 5 m above ground level. By having five systems the variance between these can be directly determined for the same conditions to assess instrumental and processing uncertainty. All systems used a Gill R3 sonic anemometer; 4 used a Vaisala

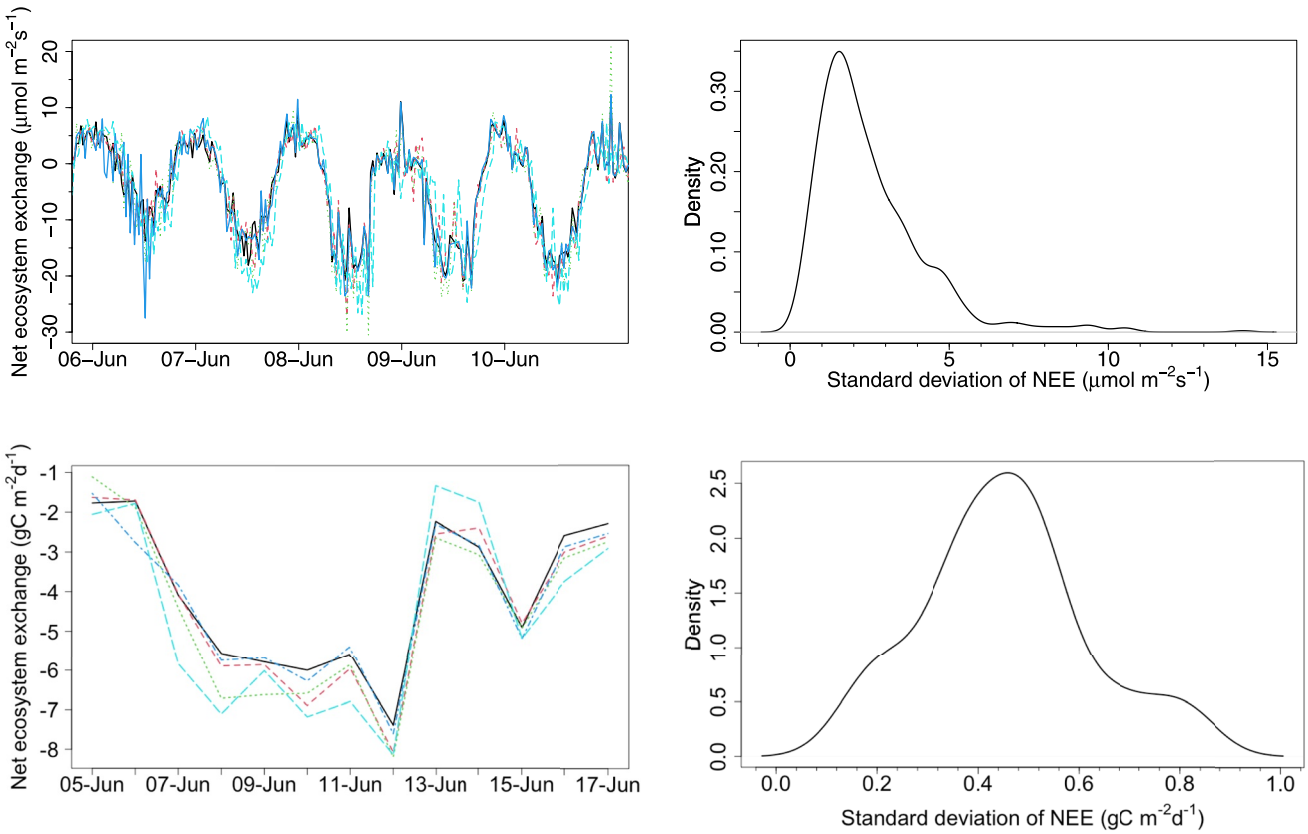

Figure 6. Comparison of five simultaneous eddy flux measurement systems (multiple time series during 2015 in left panels) and their variability (density of standard deviation across all systems for each time interval, right panels) over 30 min (top panels) and daily (bottom panels) time integrals. The five eddy systems were all installed in the same field, sown with grass, and all within 10 m of each other. For 13 days there were continuous 30-min time samples available for comparison (except 9-June, only 43 samples available for comparison) or for summing to produce a daily estimate. Only 5 days are shown in the top left panel for clarity. Density plots used a Gaussian kernel, with bandwidth selected based on "Silverman's rule of thumb".

GMP343 CO_2 sensor and a Honeywell HIH-4000 relative humidity sensor, while one used an LI-COR LI-7500 (Hill et al., 2017). Fluxes were calculated for 30-min periods using EdiRe (version 1.5.0.50). There were 13 days during June 2015 with almost continuous and simultaneous measurements by all sensor systems. The pasture was cut on 13 May, and LAI was tracked using an LAI-2000 (LiCor Inc), growing from 3 to 4 during the study period from 5 to 17 June. The deviations of the 30 min time series and the daily time series from aggregated data were calculated, and their distributions were tested for normality using a Shapiro-Wilk test.

There was broadly a good agreement between all the flux systems in their estimates of NEE at 30 min and daily time steps (Figure 6). Half-hourly flux estimates from the different systems deviated from the mean by 2 μmol m^{-2} s^{-1} (median value). The 30-min uncertainty distribution deviated significantly from normal ($P < 0.001$), being clearly log-normal with a high tail (Figure 6). For the daily aggregate NEE estimates, the median value was 0.45 gC m^{-2} d^{-1}, and the error distribution was not significantly different from normal (Shapiro-Wilk test $p = 0.93$). These results are close to the results of space-for-time uncertainty estimates at Harvard Forest, which reported a daily standard deviation of 0.58 g C m^{-2} d^{-1} (Hill et al., 2012) for NEE.

Some of the closest comparable studies are those of Katul et al. (1999) and Oren et al. (2006). They used seven towers separated by at least 100 m, so set out to deliberately measure different, non-overlapping footprints as far as possible, making the assumption that the surface flux would be the same. Both are valid but contrasting approaches, the difference amounting to the perspective on which is the biggest cause of variability in the measured flux: spatial variation in the surface flux (because of different ecophysiological properties in the different regions); or spatial variation in the turbulent flux (because of differences in the turbulence in the different regions).

An obvious extension of this work would be to run parallel systems in the longer term to see the cumulative effect on annual budgets. It is currently not clear to what extent the uncertainties measured here are random, and so will cancel out over time, or systematic, and so accumulate over time. Similarly, we do not know whether these uncertainties are relative or absolute quantities. With the substantial CO_2 fluxes over the summer, these uncertainties are relatively small; if the absolute magnitude remained constant, this would have a large effect on the small fluxes over the winter. Current processing methods are improving in their handling of uncertainty, for example, by incorporating a small ensemble of processing variants (such as different threshold values for friction velocity filtering, Pastorello et al. [2020]). However, this is still some way from a comprehensive treatment of systematic uncertainty, and this provides the rationale for empirically measuring it here.

The value of EC uncertainty estimates like these is clear. Uncertainties limit the capacity of EC to detect C sources and sinks, which are driven by small differences between large input and output fluxes. These uncertainties are also important to set the weighting of EC observations in the calibration of C process models. Bayesian calibration methods weigh the importance of observational data constraints in model fitting on the basis of measurement uncertainty. Bayesian calibrations propagate measurement uncertainty into model predictions, for example, by finding an ensemble of model parameters that produce estimates of NEE consistent with observations and their uncertainty. Eddy covariance data have been used successfully across the UK to produce a more robust calibration of C cycle models. The calibrated models then propagate the flux uncertainty from localized and incomplete data sets at a few sites into complete and regional assessments of C cycling (Myrgiotis et al., 2020; Revill et al., 2016; Smallman et al., 2017).

7. Quantifying Aggregation Error in Spatial Upscaling

Chamber and micrometeorological measurements are inevitably limited in their spatial extent, and large-scale models will almost always use grid cells that are much larger. We next consider the problem that arises when we move between these scales, usually by aggregating input variables to represent the grid-cell mean, ignoring the smaller scale variability.

7.1. Analytical Approach

Typically, we derive a model of the GHG flux based on small-scale measurements as a function of input variables x, measured at the corresponding scale. We want to apply this to a larger scale, usually using the mean value of x, averaged over a grid cell, region, or longer time period. However, the output of a model f at the larger scale, with spatially varying inputs x should be calculated as $\int_{-\infty}^{+\infty} f(x)\, p(x)\, dx$, where $p(x)$ is the probability distribution of input values in the region. Van Oijen et al. (2017) denoted this integral more briefly as $E[f(x)]$, using the expectation operator $E[]$, and derived the analysis that follows. Whenever models are run with the averaged inputs, what is being calculated is $f(E[x])$ rather than $E[f(x)]$. If the models are nonlinear, this results in an upscaling error δ defined as:

$$\Delta = f(E[x]) - E[f(x)] \tag{11}$$

When models are applied to large regions, the domain is typically subdivided using a spatial grid, with each grid cell covering an area of tens or hundreds of square kilometers. In contrast, the models themselves tend to be based on observations made at much finer scales such as flux chambers, eddy covariance footprints, individual crop fields, or forest stands. The challenge here is in using a small-scale model to estimate GHG fluxes in a large grid cell, accounting for within-cell spatial heterogeneity, and the error this produces.

In the GHGEF program, Van Oijen et al. (2017) derived a method for characterizing this error, based on a multivariate Taylor-expansion approach, a development of earlier work in crop and forest modeling (Band et al., 1991; Bresler & Dagan, 1988; Rastetter et al., 1992). The approach is to estimate Δ and correct for it. For functions of one variable, we can use the formula derived by applying the expectation operator to the second-order Taylor expansion of $f(x)$:

$$E[f(x)] \approx f(E[x]) + \frac{1}{2}\mathrm{Var}[x] f^{(2)}(E[x]), \tag{12}$$

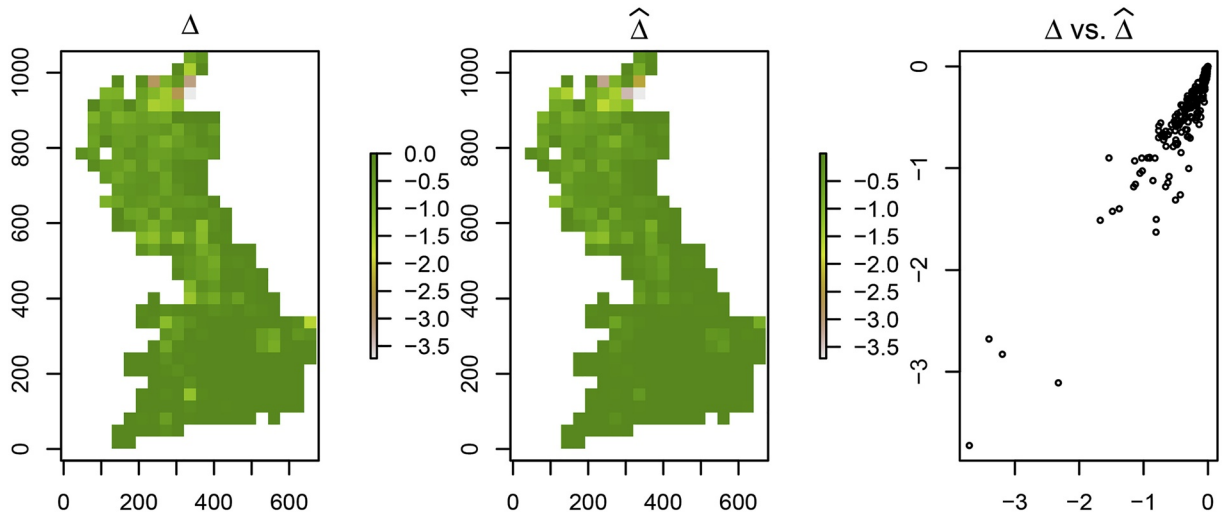

Figure 7. Left panel: The true upscaling error Δ for the methane flux model is calculated by subtracting the correctly upscaled model results (using high-resolution input data) from the incorrectly upscaled (using aggregated (mean) input data), and is mapped. Units are nmol CH_4 m^{-2} s^{-1} in all cases. Middle panel: The approximated upscaling errors that were predicted by the $\widehat{\Delta}$-formula applied to the inputs and their (co)variances. Right panel: The quality of the error-prediction can be evaluated from the scatterplot of Δ (horizontal axis) versus $\widehat{\Delta}$ (vertical axis).

where $f^{(2)}$ $(E[x])$ is the second derivative of f evaluated at the mean of x, and Var$[x]$ is the variance of x within the region. Combining this formula with Equation 11 gives us an approximate formula for the upscaling error of models with one input variable:

$$\widehat{\Delta} = -\frac{1}{2}\mathrm{Var}[x] f^{(2)}(E[x]), \tag{13}$$

where we use the ˆ-symbol to indicate that the formula provides an estimator for Δ, not Δ itself. For functions of multiple input variables where x is a vector, multivariate Taylor expansion yields:

$$\widehat{\Delta} = -\frac{1}{2}\mathrm{tr}(S\,H), \tag{14}$$

where S is the variance-covariance matrix of x, H is the Hessian matrix of second order partial derivatives of $f(x)$, and tr denotes the trace (sum of diagonal elements) of the matrix product $S\,H$. For simple models, analysis shows that the formula is exact (i.e., $\widehat{\Delta} = \Delta$), allowing full correction of model upscaling errors. In other cases, the formula provides an approximation.

We demonstrated the application of this approach to example models (Van Oijen et al., 2017) of methane flux (Levy et al., 2012), ammonia and nitrous oxide flux (Flechard et al., 2007). Using high-resolution data, we could calculate the true model output $E[f(x)]$. After averaging the input data in 32-km grid cells, we re-evaluated the model using this, corresponding to $f(E[x])$, and by difference, we calculated the aggregation error Δ. Using the Taylor-expansion approach described above, we then calculated our approximation of this error $\widehat{\Delta}$, for verification against the known values of Δ. Depending on the model, aggregation error could be substantial, ranging from −3 to −48%. The error varied spatially (Figure 7), depending on the range of the input variables, their sub-grid variance, and the non-linearity of the model. The $\widehat{\Delta}$ formula gave reasonable approximations to the true error (Figure 7, right-hand panel), correcting model output to within 2% to −9% of the true values.

For nonlinear models, the effects of spatial upscaling need to be accounted for, and the $\widehat{\Delta}$ formula described here is a generally applicable means to do this. The approach can be applied to more complex, high-dimensional process-based models, but the exploration of the accuracy of the approximation is needed.

Table 1
Mean Annual C-Budgets for Great Britain for Each Target Resolution

Resolution	GPP	Reco	NEE	Forest loss	NBE
5 km	13.3 (10.1/16.0)	12.8 (8.6/18.3)	−0.39 (−4.1/4.4)	0.03 (0.01/0.06)	−0.39 (−4.1/4.4)
28 km	13.2 (10.0/15.9)	12.7 (8.5/18.2)	−0.37 (−4.1/4.5)	0.03 (0.01/0.06)	−0.37 (−4.1/4.5)
56 km	13.1 (9.9/16.1)	12.6 (8.5/18.2)	−0.41 (−4.1/4.3)	0.02 (0.01/0.04)	−0.41 (−4.1/4.3)
111 km	14.3 (11.4/16.6)	13.8 (9.9/19)	−0.32 (−3.7/4.2)	0.01 (0.005/0.02)	−0.32 (−3.7/4.2)

Note. Budget terms are presented as the mean annual flux in MgC ha^{-1} yr^{-1}. Budget terms presented are gross primary productivity (GPP), ecosystem respiration (Reco), net ecosystem exchange of carbon (NEE = Reco − GPP) C loss due to forest removal (Forest loss) and net biome exchange of carbon (NBE = NEE + Fire loss). The 95% confidence interval of each term is presented in parenthesis. Fire emissions were low and so are not included in the summary.

7.2. An Approach Using Earth Observation and Model-Data Fusion

A more direct approach to quantifying aggregation error is to provide inputs at increasingly fine scales. By doing this, we can empirically measure the change in the output with the change in the scale of the inputs, and thus the sensitivity and corresponding aggregation error. Earth observation data (and other relevant mapped products) provide a means to do this, as they are available over wide domains at increasingly fine scales. To this end, we conducted four model-data fusion analyses at differing spatial resolutions of Great Britain's (GB) terrestrial carbon cycle at a monthly time step for an 18 year period (2001–2018) using the CARbon DAta Model fraMework (CARDAMOM, Bloom et al. [2016]). CARDAMOM uses a Bayesian approach within an Adaptive Proposal - Markov Chain Monte Carlo (AP-MHMCMC, Haario et al. [2001]; Roberts & Rosenthal [2009]) to estimate location (i.e., pixel) specific ensembles of parameters for an intermediate complexity model of the terrestrial C-cycle (DALEC Bloom & Williams [2015]). The parameter ensembles are consistent with observational constraints, their uncertainties, model structure, meteorology, and disturbance (fire & forest loss). From these parameter ensembles, we are able to directly estimate at pixel level the uncertainty of DALECs C-cycle simulation of terrestrial fluxes and stocks. CARDAMOM analyses were conducted at 4 spatial resolutions, 111 × 111, 56 × 56, 28 × 28, and 5 × 5 km, all at monthly temporal resolution. Observational constraints assimilated by CARDAMOM are monthly time series information on leaf area index (LAI), a single estimate of above ground biomass (AGB), and soil C stocks. LAI is extracted from the 1 × 1 km, 8 days product from the Copernicus Service Information (2020). A single per-pixel estimate of AGB for 2017 and its uncertainty is drawn from ESA's 1 × 1 km CCI Biomass product (Santoro, 2021). From this AGB estimate, we derive the total woody biomass (which corresponds with DALECs model structure) following Saatchi et al. (2011). A single per-pixel estimate of soil carbon stock is extracted from the SoilGrids database (Hengl et al., 2017). Meteorological drivers are drawn from the ERA5 reanalysis (Hersbach et al., 2020).

Our analyses all estimate that GB's terrestrial ecosystems were a net sink of carbon with a net biome exchange (NBE) of −6.7 to −10 TgC yr^{-1} (−0.32 to −0.41 MgC ha yr^{-1}) between 2001 and 2017 (Table 1). Spatial resolution has a substantial impact on the magnitude and spatial patterning estimated by the CARDAMOM analyses (Table 1, Figure 8). For example, at the coarsest resolution of the model grid (111 × 111 km) NBE is near neutral across much of GB except the far north east and south west. In contrast, the highest resolution (5 × 5 km) shows fine scale variation across the whole of GB. Moreover, the range of C flux magnitudes estimates is smaller in the coarser resolution analyses due to aggregation of sub-grid variability. The distributions of pixel-level mean annual fluxes progressively converge toward that estimated by our finest spatial resolution analysis (Figures 8 and 9).

The estimates of the gross biological fluxes (i.e., GPP and Reco) are relatively insensitive to spatial resolution between the 5, 28, and 56 km analyses. The mean annual flux estimate of GPP and Reco in the 5, 28, and 56 km resolution analyses vary by less than 0.2 MgC ha yr^{-1} (<2%; Table 1). Estimation of emissions due to disturbance, both fire and forest cover loss, are progressively underestimated at coarser spatial resolutions (Table 1) and show varied time series dynamics (Figure 9). The mean pixel-level forest loss estimates plateaus between 5 and 28 km (Table 1). However, there remains disagreement in the temporal interannual variability and magnitude of forest loss estimates, particularly after 2010 (Figure 9). Emissions due to fire have not converged in terms of mean annual emissions (Table 1), or temporal dynamics and magnitude (Figure 9). Overall, these results suggest

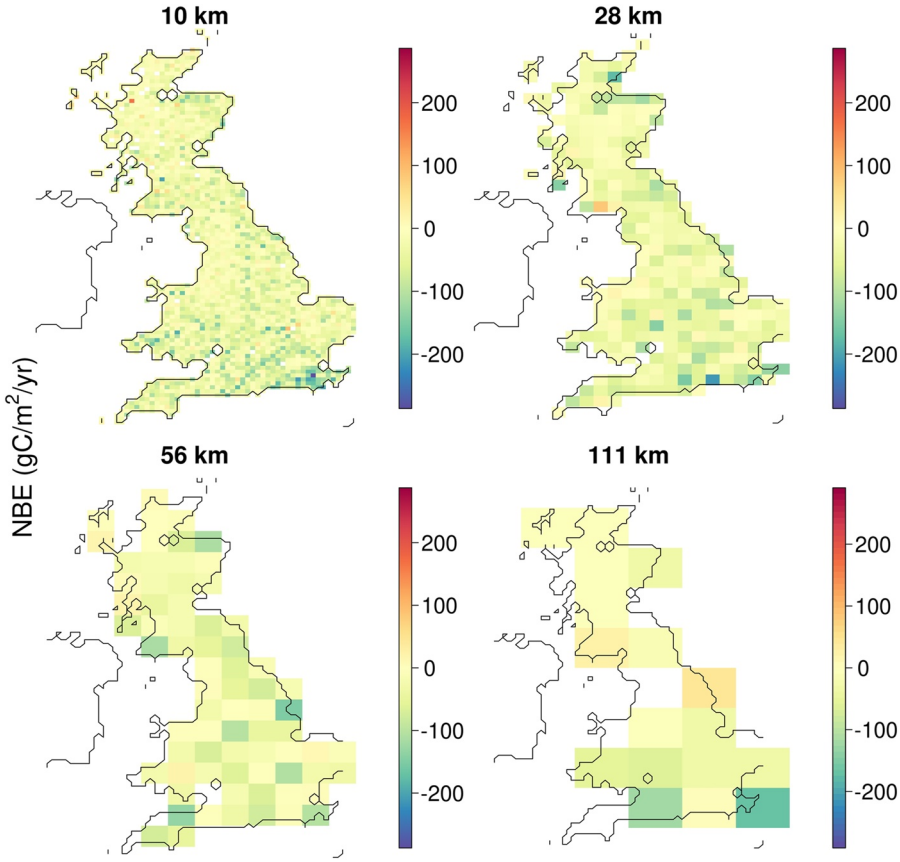

Figure 8. Net Biome Exchange (NBE = −GPP + Reco + Fire) estimated by 4 CARDAMOM analyses at a range of spatial resolutions. A negative value indicates a net uptake of carbon.

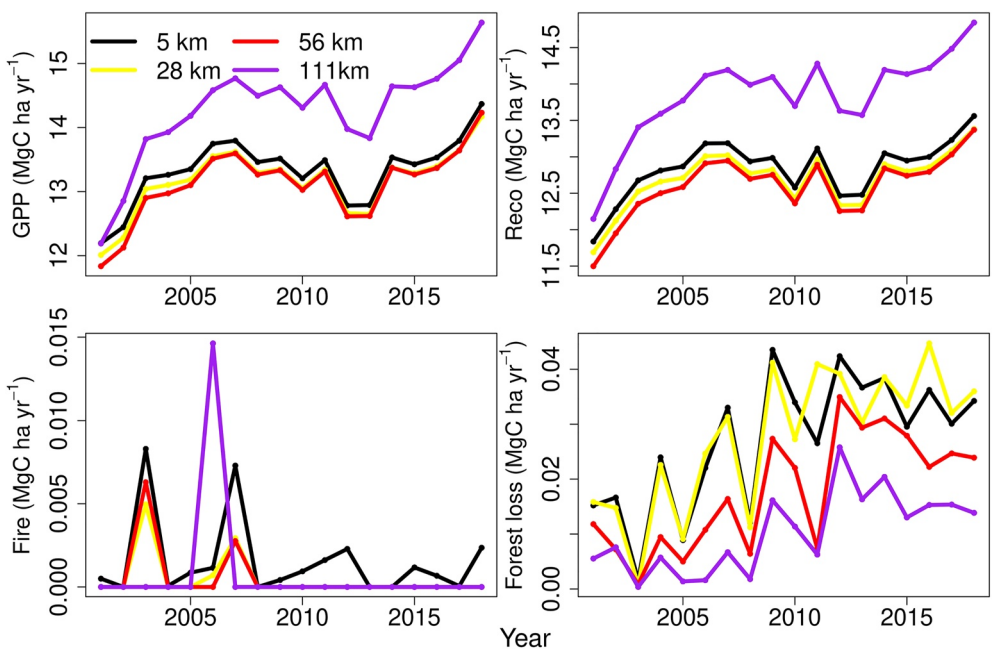

Figure 9. Time series of GB wide mean Gross Primary Productivity (GPP), ecosystem respiration (Reco), fire and forest loss estimated by 4 CARDAMOM analyses at a range of spatial resolutions.

that there remains substantial sub-grid scale disturbance information missing from these analyses, even at 5 km spatial resolution.

8. Discussion and Conclusions

We described above some of the major challenges in accurately estimating net fluxes of GHGs at the regional or national scale. These revolve around the difficulties of extrapolating small-scale, short-term observations and models. Here, we consider some commonalities across these individual studies and their findings. Perhaps the most important point is to recognize the basic upscaling problem: that we cannot observe the property of interest, and that all upscaling involves a model of some form. In many studies, the upscaling model is implicit (and the uncertainties thereby ignored) but this needs to be made explicit if the associated uncertainties are to be represented.

Returning to Equation 1, we need to be aware that there is uncertainty in the input variables x over the large-scale domain, and in how the parameters θ derived at a small scale map on to those appropriate to the large scale. In Section 3, a naïve approach would be to assume that the sample means from the flux chambers could be equated to the whole-field domain mean. However, depending on the spatial pattern across the domain, the samples may give a better or worse approximation to the true domain mean, but this spatial pattern is not known. The Bayesian approach allows us to quantify this uncertainty, by representing the distribution of possible spatial patterns that are consistent with the data. In Section 4.2, the naïve approach is to assume that the temporal pattern is based purely on the time series of sample means from the flux chambers. In fact, the measurements give a poorer constraint because of their strongly skewed spatial distribution, but we have expectations about the temporal pattern from prior knowledge of the processes. Again, the Bayesian approach allows us to combine these into the large-scale estimate (the cumulative flux). In Section 5.1, the manual static chamber method would typically provide samples over the standard working day, and equate this with the daily mean. Any effects of diurnal cycles (Keane et al., 2017) are implicitly ignored. Having a continuous hourly time series allows us to model such effects, and account for any bias that would be introduced into the long-term mean. In Section 7.1, we show an analytical method that quantifies the consequences of ignoring the upscaling steps and thereby provides a means for correction, where the necessary variances and covariances can be estimated.

Another point is that it is important to recognize that many of the things we consider to be measurements involve a model; what we actually measure is an indirect variable for example, mixing ratios, radiance at a satellite sensor, soil carbon, etc. At a lower level of detail, what we really measure is often a voltage in an instrument system, related to the indirect variable via a calibration model. In the case of eddy covariance, we do not measure the flux, but a time series of mixing ratios and w, and we have to infer their true covariance, given a range of assumptions and estimated corrections. In Section 5.2, the standard methodology would be to assume that using covariance maximization each half-hour would produce the correct flux. We need to explicitly represent the uncertainty in the true time series of time lags, and propagate the effect of this on the calculated flux, if we are to estimate the correct long-term mean and its uncertainty. Again, this needs to be made explicit if the associated uncertainties are to be represented when the measurements are extrapolated in time and/or space. In some chamber measurement protocols, only a single gas sample is used to estimate the flux, inferring $d\chi/dt_0$ by comparison with ambient air samples (Chadwick et al., 2014). This precludes any estimate of measurement error or assessment of non-linearity, so foregoes the option of quantifying uncertainty in the upscaled results.

A further point is that we need to consider the likely sources of uncertainty in the upscaling process, in terms of f, θ and x, and their different nature. These issues will be different for different GHGs because of the different processes involved. In some cases, the underlying model may be relatively linear, but the distribution of the inputs is poorly known. For example, the ecosystem-scale response of photosynthesis to radiation at a daily time scale is close to linear, and the light-use efficiency parameters vary within a quite narrow range governed by C3 biochemistry. However, the spatial distribution of radiation incident at the surface over large domains is rather poorly known on a day-to-day basis because of the vagaries of cloud cover and the sparsity of radiation measurements on the ground. The latter may therefore be the factor limiting the accuracy of large-scale predictions. By contrast, in other cases, the underlying model may be non-linear and the parameters uncertain, whilst the distribution of the inputs is well constrained. For example, the response of methane emissions to soil moisture is highly non-linear, and rather variable among vegetation types, but soil moisture itself is relatively predictable

from rainfall, topography, and soil physics. These different factors need to be accounted for when considering how best to reduce the uncertainty in upscaled estimates.

With advances in computing power, running models of GHG fluxes over large domains at relatively high resolution is increasingly feasible. For example, we can now routinely run models over the UK at 1-km resolution (Robinson et al., 2016), requiring 240,000 grid cells. In moving from the 1 km^2 scale to the national scale, the issues of model upscaling do not apply in the same way, so long as the model has been correctly scaled from the original observations to the 1-km^2 scale. This is because the variance over the whole domain is represented, albeit as 1-km^2 means; that is, we evaluate E[f(x)], so $\Delta = 0$, so the total flux is simply the sum of all the grid cells. In moving between these scales, the main problem encountered is in estimating the input variables accurately over this domain. This is not intrinsically an upscaling problem, in that it is not fundamentally linked to changing the scale at which the model is applied, but presents a serious sampling problem in accurately estimating over a large region. Because of the large size of the domain, our information is often incomplete or based on proxies (e.g., NDVI) Disney et al. (2016). This is an area where data availability is rapidly improving, with new and improved satellite products becoming available, as well as other new technologies - improved laser-based spectroscopic methods, and unmanned airborne vehicles.

Lastly, we provide some considerations for future work on upscaling. A key point is simply to recognize that the scaling problem exists and that some attempt is made to deal with it explicitly, rather than hide it in implicit assumptions. The field of geostatistics has dealt with similar problems, of estimating over large scales from restricted sampling, over several decades, but the techniques could be more widely applied in biogeochemistry and process-based modeling. We note that the same issues occur in the time domain as well as the spatial domain, and this should be considered also.

A common theme throughout is the adoption of Bayesian principles for coherently combining multiple information sources, which allows us to make better use of the available data. We expect this trend to continue and expand into the machine learning domain.

The linearity of the underlying model is an important factor in determining the difficulty of upscaling, and this is very problem-specific. However, some analysis of the processes and the form of the model can be done to assess the scope for alleviating the difficulties. For example, the response of plant canopies to radiation is non-linear over short (half-hourly) time scales because of light saturation, but relatively linear at the daily time scale (Ruimy et al., 1995). A judicious choice of scales to consider may be helpful in such cases.

Using the method from Section 7.1, the extent of upscaling aggregation error can be estimated, given estimated variability in the small-scale inputs. The latter may not be known exactly, but some proxies may be available which will give information on their distribution. For example, GHG emissions may be sensitive to soil moisture, and its fine-scale spatial variability may not be known in absolute terms. However, the topography is often well described, in the form of a high-resolution digital elevation model. This can be used to compute a topographic wetness index (Beven & Kirkby, 1979; Marthews, 2015), which may give a reasonable proxy for the distribution of soil moisture. Similarly, remotely-sensed data can provide information on the distribution and variance of proxy variables over a wide domain. The errors and biases in the satellite retrieval process need to be considered when assessing how good a proxy they provide.

Even if the small-scale distribution is unknown and no proxies are available, the approaches from Section 7 can be applied. By using a range of variances and covariances in the analytical approach, the likely range in aggregation error can be assessed, and the sensitivity of the upscaling process to this examined. Or more empirically, by using simulation with a series of aggregated input data, we can achieve the same thing, to show the effect of variance at different scales.

The uncertainties present in measurements are important but often difficult to estimate rigorously and should be empirically measured where possible. Exercises like the replicated eddy covariance systems in Section 6 are worth including in future studies. If run over the longer term, this would be particularly useful in separating aleatoric and epistemic uncertainty - that is, whether the uncertainty is effectively random, or due to imperfectly accounting for real effects for example, related to spatial heterogeneity and differences in sensor placement.

Although direct measurement at the large scale of interest is not possible (hence the need for upscaling), it is worth exploring whether any pertinent information is available at this scale. As example: airborne flux measurements,

while often only available on a campaign basis for limited flight lines, provide a means of testing upscaling procedures. Rivers provide natural integration over their catchment areas, and outflow data (in relation to rainfall) can provide indirect information on large-scale evapotranspiration, with some bearing on other gas fluxes (Martínez-de la Torre et al., 2019). Tall-tower measurements of GHG mixing ratios can be used in inverse atmospheric modeling to infer regional-scale GHG fluxes (Manning et al., 2011; White et al., 2018). The Bayesian approach allows these data to be used in a consistent framework, so long as the underlying model makes predictions of these measured quantities.

In addition to the upscaling issues considered here, it is possible that new phenomena may arise at the larger scale because of feedback in the system, which are not apparent at the small scale. In this case, parameters θ, which we assumed to be fixed at the small scale, become variables x at the larger scale, so the form of the model changes with scale. For example, evapotranspiration from an individual leaf is strongly controlled by stomatal conductance. However, because of the effect of regional-scale evapotranspiration on the vapor pressure deficit of the air in the boundary layer, regional-scale evapotranspiration is more strongly controlled by radiation input. This is a serious issue with water vapor fluxes, but less so for GHG fluxes themselves, because the magnitude of such feedback is generally much smaller. However, the possibilities should be considered in any given upscaling problem.

On a philosophical note, Karl Popper observed that the art of science is in "discerning what we may with advantage omit". Clearly, we cannot try to represent every single component in the measurement system and ecosystem when considering regional-scale domains. The judgment is in discerning which parts matter and which do not, and this may not be obvious a priori, and require numerical simulation experimentation. Lastly, we note that some of the major challenges in science can be seen as upscaling problems. Major advances in physics (such as the modern understanding of gravity) came about because of discrepancies when applying physical models derived on the spatial scale we are familiar with to the much larger scales of astronomy. Upscaling should not be seen as a minor inconvenience getting in the way of what we actually want to predict, but as an important step in the scientific process.

Data Availability Statement

Data sets and code used in the study are provided in a GitHub repository associated with the paper via https://doi.org/10.5281/zenodo.6402307 (P. Levy, 2022). Other data for the program are being archived at the UK Environmental Information Data Centre (https://eidc.ac.uk/) and the British Atmospheric Data Centre (https://catalogue.ceda.ac.uk/uuid/c117a2e6f451405393cac1c6fbf8f7a3).

Acknowledgments
We thank two anonymous reviewers for their comments and suggestions which greatly improved the manuscript. The work was funded by the UK Natural Environment Research Council, grant references: NE/K002481/1, NE/K002619/1, and NE/K002538/1.

References

Aubinet, M., Vesala, T., & Papale, D. (Eds.). (2012). *Eddy covariance: A Practical Guide to measurement and data analysis*. Springer Netherlands.

Band, L. E., Peterson, D. L., Running, S. W., Coughlan, J., Lammers, R., Dungan, J., & Nemani, R. (1991). Forest ecosystem processes at the watershed scale: Basis for distributed simulation. *Ecological Modelling*, *56*, 171–196. https://doi.org/10.1016/0304-3800(91)90199-B

Beven, K., & Kirkby, P. (1979). *A physically based variable contribution area model of basin hydrology*.

Bloom, A. A., Exbrayat, J.-F., van der Velde, I. R., Feng, L., & Williams, M. (2016). The decadal state of the terrestrial carbon cycle: Global retrievals of terrestrial carbon allocation, pools, and residence times. *Proceedings of the National Academy of Sciences*, *113*(5), 1285–1290. https://doi.org/10.1073/pnas.1515160113

Bloom, A. A., & Williams, M. (2015). Constraining ecosystem carbon dynamics in a data-limited world: Integrating ecological "common sense" in a model–data fusion framework. *Biogeosciences*, *12*(5), 1299–1315. https://doi.org/10.5194/bg-12-1299-2015

Bresler, E., & Dagan, G. (1988). Variability of yield of an irrigated crop and its causes: 1. Statement of the problem and methodology. *Water Resources Research*, *24*(3), 381–387. https://doi.org/10.1029/WR024i003p00381

Burba, G. (2013). *Eddy covariance method*. Li-COR Biogeosciences.

Chadwick, D. R., Cardenas, L., Misselbrook, T. H., Smith, K. A., Rees, R. M., Watson, C. J., et al. (2014). Optimizing chamber methods for measuring nitrous oxide emissions from plot-based agricultural experiments. *European Journal of Soil Science*, *65*(2), 295–307. https://doi.org/10.1111/ejss.12117

Cowan, N. J., Famulari, D., Levy, P. E., Anderson, M., Bell, M. J., Rees, R. M., et al. (2014). An improved method for measuring soil N2O fluxes using a quantum cascade laser with a dynamic chamber: Dynamic chamber method. *European Journal of Soil Science*, *65*(5), 643–652. https://doi.org/10.1111/ejss.12168

Cowan, N. J., Levy, P. E., Drewer, J., Carswell, A., Shaw, R., Simmons, I., et al. (2019). Application of Bayesian statistics to estimate nitrous oxide emission factors of three nitrogen fertilisers on UK grasslands. *Environment International*, *128*, 362–370. https://doi.org/10.1016/j.envint.2019.04.054

Cowan, N. J., Levy, P. E., Famulari, D., Anderson, M., Drewer, J., Carozzi, M., et al. (2016). The influence of tillage on N2O fluxes from an intensively managed grazed grassland in Scotland. *Biogeosciences*, *13*(16), 4811–4821. https://doi.org/10.5194/bg-13-4811-2016

Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, *22*(3), 239–252. https://doi.org/10.1007/BF00889887

Desjardins, R. L., MacPherson, J. I., Mahrt, L., Schuepp, P., Pattey, E., Neumann, H., et al. (1997). Scaling up flux measurements for the boreal forest using aircraft-tower combinations. *Journal of Geophysical Research*, *102*(D24), 29125–29133. https://doi.org/10.1029/97JD00278

Disney, M., Muller, J.-P., Kharbouche, S., Kaminski, T., Voßbeck, M., Lewis, P., & Pinty, B. (2016). A new global fAPAR and LAI dataset derived from optimal Albedo estimates: Comparison with MODIS products. *Remote Sensing*, *8*(4), 275. https://doi.org/10.3390/rs8040275

Finkelstein, P. L., & Sims, P. F. (2001). Sampling error in eddy correlation flux measurements. *Journal of Geophysical Research*, *106*(D4), 3503–3509. https://doi.org/10.1029/2000JD900731

Flechard, C. R., Ambus, P., Skiba, U., Rees, R. M., Hensen, A., Van Amstel, A., et al. (2007). Effects of climate and management intensity on nitrous oxide emissions in grassland systems across Europe. *Agriculture, Ecosystems & Environment*, *121*(1), 135–152. https://doi.org/10.1016/j.agee.2006.12.024

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd ed.). Chapman and Hall/CRC.

Gifford, R. (1994). The global carbon cycle: A viewpoint on the missing sink. *Australian Journal of Plant Physiology*, *21*(1), 1–15. https://doi.org/10.1071/PP9940001

Gilhespy, S. L., Anthony, S., Cardenas, L., Chadwick, D., del Prado, A., Li, C., et al. (2014). First 20 years of DNDC (DeNitrification DeComposition): Model evolution. *Ecological Modelling*, *292*, 51–62. https://doi.org/10.1016/j.ecolmodel.2014.09.004

Gioli, B., Miglietta, F., De Martino, B., Hutjes, R. W. A., Dolman, H. A. J., Lindroth, A., et al. (2004). Comparison between tower and aircraft-based eddy covariance fluxes in five European regions. *Agricultural and Forest Meteorology*, *127*(1), 1–16. https://doi.org/10.1016/j.agrformet.2004.08.004

Haario, H., Saksman, E., & Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, *7*(2), 223–242. https://doi.org/10.2307/3318737

Haas, E., Klatt, S., Fröhlich, A., Kraft, P., Werner, C., Kiese, R., et al. (2013). LandscapeDNDC: A process model for simulation of biosphere–atmosphere–hydrosphere exchange processes at site and regional scale. *Landscape Ecology*, *28*(4), 615–636. https://doi.org/10.1007/s10980-012-9772-x

Haszpra, L., Hidy, D., Taligás, T., & Barcza, Z. (2018). First results of tall tower based nitrous oxide flux monitoring over an agricultural region in Central Europe. *Atmospheric Environment*, *176*, 240–251. https://doi.org/10.1016/j.atmosenv.2017.12.035

Hengl, T., de Jesus, J. M., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Hill, T., Chocholek, M., & Clement, R. (2017). The case for increasing the statistical power of eddy covariance ecosystem studies: Why, where and how? *Global Change Biology*, *23*(6), 2154–2165. https://doi.org/10.1111/gcb.13547

Hill, T., Ryan, E., & Williams, M. (2012). The use of CO2 flux time series for parameter and carbon stock estimation in carbon cycle research [Journal Article]. *Global Change Biology*, *18*, 179–193. https://doi.org/10.1111/j.1365-2486.2011.02511.x

Hollinger, D. Y., & Richardson, A. (2005). Uncertainty in eddy covariance measurements and its application to physiological models. *Tree Physiology*, *25*, 873–885. https://doi.org/10.1093/treephys/25.7.873

Hutchinson, G. L., & Mosier, A. R. (1981). Improved soil cover method for field measurement of nitrous oxide Fluxes1. *Soil Science Society of America Journal*, *45*(2), 311–316. https://doi.org/10.2136/sssaj1981.03615995004500020017x

Ibrom, A., Dellwik, E., Flyvbjerg, H., Jensen, N. O., & Pilegaard, K. (2007). Strong low-pass filtering effects on water vapour flux measurements with closed-path eddy correlation systems. *Agricultural and Forest Meteorology*, *147*(3–4), 140–156. https://doi.org/10.1016/j.agrformet.2007.07.007

IPCC. (2013). *Climate change 2013: The physical science basis. Contribution of working Group I to the Fifth assessment Report of the Intergovernmental panel on climate change*. Cambridge University Press.

Katul, G., Hsieh, C.-I., Bowling, D., Clark, K., Shurpali, N., Turnipseed, A., et al. (1999). Spatial variability of turbulent fluxes in the roughness Sublayer of an even-aged pine forest. *Boundary-Layer Meteorology*, *93*(1), 1–28. https://doi.org/10.1023/A:1002079602069

Keane, B. J., Ineson, P., Vallack, H. W., Blei, E., Bentley, M., Howarth, S., & Toet, S. (2017). Greenhouse gas emissions from the energy crop oilseed rape (Brassica napus); the role of photosynthetically active radiation in diurnal N2O flux variation. *GCB Bioenergy*, *10*(5), 306–319. https://doi.org/10.1111/gcbb.12491

Kormann, R., & Meixner, F. X. (2001). An analytical footprint model for non-neutral stratification. *Boundary-Layer Meteorology*, *99*(2), 207–224. https://doi.org/10.1023/a:1018991015119

Kroon, P. S., Hensen, A., Jonker, H. J. J., Ouwersloot, H. G., Vermeulen, A. T., & Bosveld, F. C. (2010). Uncertainties in eddy covariance flux measurements assessed from CH4 and N2O observations. *Agricultural and Forest Meteorology*, *150*(6), 806–816. https://doi.org/10.1016/j.agrformet.2009.08.008

Langford, B., Acton, W., Ammann, C., Valach, A., & Nemitz, E. (2015). Eddy-covariance data with low signal-to-noise ratio: Time-lag determination, uncertainties and limit of detection. *Atmospheric Measurement Techniques*, *8*(10), 4197–4213. https://doi.org/10.5194/amt-8-4197-2015

Leclerc, M. Y., & Foken, T. (2014). *Footprints in Micrometeorology and ecology*. Springer-Verlag.

Lee, X., Massman, W., & Law, B. (2006). *Handbook of Micrometeorology: A Guide for surface flux measurement and analysis*. Springer Science & Business Media.

Leip, A., Skiba, U., Vermeulen, A., & Thompson, R. L. (2018). A complete rethink is needed on how greenhouse gas emissions are quantified for national reporting. *Atmospheric Environment*, *174*, 237–240. https://doi.org/10.1016/j.atmosenv.2017.12.006

Levy, P. (2022). NERC-CEH/upscaling_JGR_SI. *Zenodo*. Initial release. https://doi.org/10.5281/zenodo.6402307

Levy, P. E., Burden, A., Cooper, M. D. A., Dinsmore, K. J., Drewer, J., Evans, C., et al. (2012). Methane emissions from soils: Synthesis and analysis of a large UK data set. *Global Change Biology*, *18*(5), 1657–1669. https://doi.org/10.1111/j.1365-2486.2011.02616.x

Levy, P., Drewer, J., Jammet, M., Leeson, S., Friborg, T., Skiba, U., & van Oijen, M. (2020). Inference of spatial heterogeneity in surface fluxes from eddy covariance data: A case study from a subarctic mire ecosystem. *Agricultural and Forest Meteorology*, *280*, 107783. https://doi.org/10.1016/j.agrformet.2019.107783

Levy, P. E., Cowan, N., van Oijen, M., Famulari, D., Drewer, J., & Skiba, U. (2017). Estimation of cumulative fluxes of nitrous oxide: Uncertainty in temporal upscaling and emission factors. *European Journal of Soil Science*, *68*(4), 400–411. https://doi.org/10.1111/ejss.12432

Levy, P. E., Gray, A., Leeson, S. R., Gaiawyn, J., Kelly, M. P. C., Cooper, M. D. A., et al. (2011). Quantification of uncertainty in trace gas fluxes measured by the static chamber method. *European Journal of Soil Science*, *62*(6), 811–821. https://doi.org/10.1111/j.1365-2389.2011.01403.x

Livingston, G. P., Hutchinson, G. L., & Spartalian, K. (2006). Trace gas emission in chambers. *Soil Science Society of America Journal*, *70*(5), 1459–1469. https://doi.org/10.2136/sssaj2005.0322

Mammarella, I., Werle, P., Pihlatie, M., Eugster, W., Haapanala, S., Kiese, R., et al. (2010). A case study of eddy covariance flux of N2O measured within forest ecosystems: Quality control and flux error analysis. *Biogeosciences*, *7*(2), 427–440. https://doi.org/10.5194/bg-7-427-2010

Manning, A. J., O'Doherty, S., Jones, A. R., Simmonds, P. G., & Derwent, R. G. (2011). Estimating UK methane and nitrous oxide emissions from 1990 to 2007 using an inversion modeling approach. *Journal of Geophysical Research*, *116*(D2), D02305. https://doi.org/10.1029/2010JD014763

Marthews, T. R. (2015). *High-resolution global topographic index values*. NERC Environmental Information Data Centre. https://doi.org/10.5285/6b0c4358-2bf3-4924-aa8f-793d468b92be

Martínez-de la Torre, A., Blyth, E. M., & Weedon, G. P. (2019). Using observed river flow data to improve the hydrological functioning of the JULES land surface model (vn4.3) used for regional coupled modelling in Great Britain (UKC2). *Geoscientific Model Development*, *12*(2), 765–784. https://doi.org/10.5194/gmd-12-765-2019

Matson, P. A., & Harriss, R. C. (2009). *Biogenic trace gases: Measuring emissions from soil and water*. John Wiley & Sons.

Moncrieff, J. B., Massheder, J. M., DeBruin, H., Elbers, J., Friborg, T., Heusinkveld, B., et al. (1997). A system to measure surface fluxes of momentum, sensible heat, water vapour and carbon dioxide. *Journal of Hydrology*, *189*(1–4), 589–611. https://doi.org/10.1016/s0022-1694(96)03194-0

Morin, T. H., Bohrer, G., Stefanik, K. C., Rey-Sanchez, A. C., Matheny, A. M., & Mitsch, W. J. (2017). Combining eddy-covariance and chamber measurements to determine the methane budget from a small, heterogeneous urban floodplain wetland park. *Agricultural and Forest Meteorology*, *237–238*, 160–170. https://doi.org/10.1016/j.agrformet.2017.01.022

Myrgiotis, V., Blei, E., Clement, R., Jones, S. K., Keane, B., Lee, M. A., et al. (2020). A model-data fusion approach to analyse carbon dynamics in managed grasslands [Journal Article]. *Agricultural Systems*, *184*, 102907. https://doi.org/10.1016/j.agsy.2020.102907

Myrgiotis, V., Williams, M., Rees, R. M., Smith, K. E., Thorman, R. E., & Topp, C. F. E. (2016). Model evaluation in relation to soil N2O emissions: An algorithmic method which accounts for variability in measurements and possible time lags. *Environmental Modelling & Software*, *84*(Supplement C), 251–262. https://doi.org/10.1016/j.envsoft.2016.07.002

Myrgiotis, V., Williams, M., Rees, R. M., & Topp, C. F. E. (2019). Estimating the soil N2O emission intensity of croplands in northwest Europe. *Biogeosciences*, *16*(8), 1641–1655. https://doi.org/10.5194/bg-16-1641-2019

Oren, R., Hsieh, C.-I., Stoy, P., Albertson, J., Mccarthy, H. R., Harrell, P., & Katul, G. G. (2006). Estimating the uncertainty in annual net ecosystem carbon exchange: Spatial variation in turbulent fluxes and sampling errors in eddy-covariance measurements. *Global Change Biology*, *12*(5), 883–896. https://doi.org/10.1111/j.1365-2486.2006.01131.x

Pardo-Iguzquiza, E., & Chica-Olmo, M. (2008). Geostatistics with the matern semivariogram model: A library of computer programs for inference, kriging and simulation. *Computers & Geosciences*, *34*(9), 1073–1079. https://doi.org/10.1016/j.cageo.2007.09.020

Pastorello, G., Trotta, C., Canfora, E., Chu, H., Christianson, D., Cheah, Y.-W., et al. (2020). The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. *Scientific Data*, *7*(1), 225. https://doi.org/10.1038/s41597-020-0534-3

Pedersen, A. R., Petersen, S. O., & Schelde, K. (2010). A comprehensive approach to soil-atmosphere trace-gas flux estimation with static chambers. *European Journal of Soil Science*, *61*(6), 888–902. https://doi.org/10.1111/j.1365-2389.2010.01291.x

Pitt, J. R., Le Breton, M., Allen, G., Percival, C. J., Gallagher, M. W., Bauguitte, S. J.-B., et al. (2016). The development and evaluation of airborne in situ N2O and CH4 sampling using a quantum cascade laser absorption spectrometer (QCLAS). *Atmospheric Measurement Techniques*, *9*(1), 63–77. https://doi.org/10.5194/amt-9-63-2016

Plummer, M. (2016). *Rjags: Bayesian Graphical models using MCMC*.

Polson, D., Fowler, D., Nemitz, E., Skiba, U., McDonald, A., Famulari, D., et al. (2011). Estimation of spatial apportionment of greenhouse gas emissions for the UK using boundary layer measurements and inverse modelling technique. *Atmospheric Environment*, *45*(4), 1042–1049. https://doi.org/10.1016/j.atmosenv.2010.10.011

Rannik, Ü., Sogachev, A., Foken, T., Göckede, M., Kljun, N., Leclerc, M. Y., & Vesala, T. (2012). Footprint analysis. In M. Aubinet, T. Vesala, & D. Papale (Eds.), *Eddy covariance: A Practical Guide to measurement and data analysis* (pp. 211–261). Springer Netherlands. https://doi.org/10.1007/978-94-007-2351-1_8

Rastetter, E. B., King, A. W., Cosby, B. J., Hornberger, G. M., O'Neill, R. V., & Hobbie, J. E. (1992). Aggregating fine-scale ecological knowledge to model coarser-scale Attributes of ecosystems. *Ecological Applications*, *2*(1), 55–70. https://doi.org/10.2307/1941889

Reuss-Schmidt, K., Levy, P. E., Oechel, W., Tweedie, C. E., Wilson, C. J., & Zona, D. (2019). Understanding spatial variability of methane fluxes in Arctic wetlands through footprint modelling. *Environmental Research Letters*, *14*, 125010. https://doi.org/10.1088/1748-9326/ab4d32

Revill, A., Bloom, A. A., & Williams, M. (2016). *Impacts of reduced model complexity and driver resolution on cropland ecosystem photosynthesis estimates [Journal Article]*. Field Crops Research.

Ribeiro, P. J., Jr., Diggle, P. J., Schlather, M., Bivand, R., & Ripley, B. (2020). *geoR: Analysis of geostatistical data [Manual]*.

Roberts, G. O., & Rosenthal, J. S. (2009). *Examples of adaptive MCMC*. *18*, 349–367. https://doi.org/10.1198/jcgs.2009.06134

Robinson, E., Blyth, E., Clark, D., Comyn-Platt, E., Finch, J., & Rudd, A. (2016). *Climate hydrology and ecology research Support system potential evapotranspiration Dataset for Great Britain* (pp. 1961–2015). [CHESS-PE]. https://doi.org/10.5285/8baf805d-39ce-4dac-b224-c926ada353b7

Ruimy, A., Jarvis, P. G., Baldocchi, D. D., & Saugier, B. (1995). CO_2 fluxes over plant canopies and solar radiation: A review. *Advances in Ecological Research*. In M. Begon, & A. H. Fitter (Eds.), (Vol. 26, pp. 1–68). Academic Press. https://doi.org/10.1016/S0065-2504(08)60063-X

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences*, *108*(24), 9899–9904. https://doi.org/10.1073/pnas.1019576108

Sachs, T., Giebels, M., Boike, J., & Kutzbach, L. (2010). Environmental controls on CH4 emission from polygonal tundra on the microsite scale in the Lena river delta, Siberia. *Global Change Biology*, *16*(11), 3096–3110. https://doi.org/10.1111/j.1365-2486.2010.02232.x

Sahoo, B., & Mayya, Y. (2010). Two dimensional diffusion theory of trace gas emission into soil chambers for flux measurements. *Agricultural and Forest Meteorology*, *150*(9), 1211–1224. https://doi.org/10.1016/j.agrformet.2010.05.009

Santoro, M., & Cartus, O. (2021). ESA Biomass Climate Change Initiative (Biomasscci): Global datasets of forest above-ground biomass for the years 2010, 2017 and 2018, v2. [Dataset]. *Centre for Environmental Data Analysis*.

Schmid, H. P., & Oke, T. (1990). A model to estimate the source area contributing to turbulent exchange in the surface layer over patchy terrain. *Quarterly Journal of the Royal Meteorological Society*, *116*(494), 965–988. https://doi.org/10.1002/qj.49711649409

Schrier-Uijl, A. P., Kroon, P. S., Hensen, A., Leffelaar, P. A., Berendse, F., & Veenendaal, E. M. (2010). Comparison of chamber and eddy covariance-based CO2 and CH4 emission estimates in a heterogeneous grass ecosystem on peat. *Agricultural and Forest Meteorology*, *150*(6), 825–831. https://doi.org/10.1016/j.agrformet.2009.11.007

Schuepp, P. H., Leclerc, M. Y., MacPherson, J. I., & Desjardins, R. L. (1990). Footprint prediction of scalar fluxes from analytical solutions of the diffusion equation. *Boundary-Layer Meteorology*, *50*(1–4), 355–373. https://doi.org/10.1007/BF00120530

Smallman, T., Exbrayat, J.-F., Mencuccini, M., Bloom, A., & Williams, M. (2017). Assimilation of repeated woody biomass observations constrains decadal ecosystem carbon cycle uncertainty in aggrading forests. *Journal of Geophysical Research: Biogeosciences*, *122*(3), 528–545. https://doi.org/10.1002/2016jg003520

Smith, J. U., Smith, P., Richter, G. M., Agostini, F. A., & Welham, S. J. (2003). Testing the adequacy of measured data for evaluating nitrogen turnover models by the dot-to-dot method. *European Journal of Soil Science*, *54*(1), 175–186. https://doi.org/10.1046/j.1365-2389.2003.00501.x

Smith, P., Martino, D., Cai, Z., Gwary, D., Janzen, H., Kumar, P., et al. (2008). Greenhouse gas mitigation in agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1492), 789–813. https://doi.org/10.1098/rstb.2007.2184

Smith, P., & Smith, P. (2004). Monitoring and verification of soil carbon changes under Article 3.4 of the Kyoto Protocol. *Soil Use & Management*, *20*(2), 264–270. https://doi.org/10.1079/SUM2004239

Stehfest, E., & Bouwman, L. (2006). N2O and NO emission from agricultural fields and soils under natural vegetation: Summarizing available measurement data and modeling of global annual emissions. *Nutrient Cycling in Agroecosystems*, *74*(3), 207–228. https://doi.org/10.1007/s10705-006-9000-7

Van Oijen, M., Cameron, D., Levy, P. E., & Preston, R. (2017). Correcting errors from spatial upscaling of nonlinear greenhouse gas flux models. *Environmental Modelling & Software*, *94*, 157–165. https://doi.org/10.1016/j.envsoft.2017.03.023

White, E. D., Rigby, M., Lunt, M. F., Ganesan, A. L., Manning, A. J., O'Doherty, S., et al. (2018). Quantifying the UK's Carbon Dioxide Flux: An atmospheric inverse modelling approach using a regional measurement network. *Atmospheric Chemistry and Physics Discussions*, 1–35. https://doi.org/10.5194/acp-2018-839

Xu, L., Furtaw, M. D., Madsen, R. A., Garcia, R. L., Anderson, D. J., & McDermitt, D. K. (2006). On maintaining pressure equilibrium between a soil CO_2 flux chamber and the ambient air. *Journal of Geophysical Research*, *111*(D8), D08S10. https://doi.org/10.1029/2005JD006435